THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

A fully data-driven approach to minimizing CVaR for portfolio of assets via SGLD with discontinuous updating

Citation for published version:
Sabanis, S & Zhang, Y 2020 'A fully data-driven approach to minimizing CVaR for portfolio of assets via SGLD with discontinuous updating' ArXiv. <https://arxiv.org/abs/2007.01672>

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Early version, also known as pre-print

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 05. May. 2023

arXiv:2007.01672v1 [q-fin.PM] 2 Jul 2020

A fully data-driven approach to minimizing CVaR for portfolio of assets via SGLD with discontinuous updating *

Sotirios Sabanis[1,2] and Ying Zhang[1]

[1]School of Mathematics, The University of Edinburgh, UK.
[2]The Alan Turing Institute, UK.

July 6, 2020

Abstract

A new approach in stochastic optimization via the use of stochastic gradient Langevin dynamics (SGLD) algorithms, which is a variant of stochastic gradient decent (SGD) methods, allows us to efficiently approximate global minimizers of possibly complicated, high-dimensional landscapes. With this in mind, we extend here the non-asymptotic analysis of SGLD to the case of discontinuous stochastic gradients. We are thus able to provide theoretical guarantees for the algorithm's convergence in (standard) Wasserstein distances for both convex and non-convex objective functions. We also provide explicit upper estimates of the expected excess risk associated with the approximation of global minimizers of these objective functions.

All these findings allow us to devise and present a fully data-driven approach for the optimal allocation of weights for the minimization of CVaR of portfolio of assets with complete theoretical guarantees for its performance. Numerical results illustrate our main findings.

1 Introduction

We are concerned in this article with the study of stochastic optimization problems of the form

$$\text{minimize} \quad U(\theta) := \mathbb{E}[f(\theta, X)], \tag{1}$$

where the gradient of f is discontinuous in $\theta \in \mathbb{R}^d$ and X is a random element with a smooth density. Within this framework, we highlight and solve the problem of minimizing CVaR (expected shortfall) of a portfolio of assets in terms of optimal selection of weights for individual assets as explained in Section 5.2.2. We offer theoretical guarantees for the approximate solution of the optimization problem (1) by generating a $\hat{\theta}$ such that the expected excess risk

$$\mathbb{E}[U(\hat{\theta})] - \inf_{\theta \in \mathbb{R}^d} U(\theta)$$

is minimized. To achieve this, we analyse the convergence properties of the stochastic gradient Langevin dynamics (SGLD) algorithm with discontinuous updating H, which is given by

$$\theta_0^\lambda = \theta_0, \quad \theta_{n+1}^\lambda = \theta_n^\lambda - \lambda H(\theta_n^\lambda, X_{n+1}) + \sqrt{2\beta^{-1}\lambda}\xi_{n+1}, \ n \in \mathbb{N}, \tag{2}$$

where θ_0 is an $\mathbb{R}^d$-valued random variable, $\lambda > 0$ is the stepsize, $\beta > 0$ is the so-called inverse temperature parameter, $H : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ is a measurable function satisfying $\nabla U(\theta) = \mathbb{E}[H(\theta, X_0)]$ with $(X_n)_{n\in\mathbb{N}}$ being an i.i.d. sequence, and $(\xi_n)_{n\in\mathbb{N}}$ is an independent sequence of standard d-dimensional

*This work was supported by The Alan Turing Institute for Data Science and AI under EPSRC grant EP/N510129/1. Y. Z. was supported by The Maxwell Institute Graduate School in Analysis and its Applications, a Centre for Doctoral Training funded by the UK Engineering and Physical Sciences Research Council (grant EP/L016508/01), the Scottish Funding Council, Heriot-Watt University and the University of Edinburgh.

Gaussian random variables. One recalls heere that the SGLD algorithm (2) can be viewed as a discretization of the Langevin SDE:

$$Z_0 = \theta_0, \quad dZ_t = -h(Z_t)dt + \sqrt{2\beta^{-1}}dB_t, \tag{3}$$

where $h := \nabla U$ and $(B_t)_{t\geq 0}$ represents the standard Brownian motion. Moreover, it is well-known that, under appropriate conditions, the Langevin SDE (3) admits a unique invariant measure $\pi_\beta \propto \exp(-\beta U(\theta))$ which concentrates around the minimizers of U when β is sufficiently large, , see [16] for more details.

Theoretical guarantees of the SGLD algorithm (2) to the target distribution π_β have been established in Wasserstein-2 distance under the assumptions that H is convex and (locally) Lipschitz continuous, see [1], [2], [10] and references therein. Recently, these results are considered under more generalised conditions aiming to include a wider range of practical applications. To relax the convexity condition, a dissipativity condition is proposed in [19], and the convergence result is obtained in Wasserstein-2 distance with the rate $\lambda^{5/4}n$. This is the first such result in non-convex optimization, which is then improved in the work [23] and [8]. Compared to [19], a higher rate of convergence with dependence on n is achieved in [23] following a direct analysis of the ergodicity of the overdamped Langevin Monte Carlo (LMC) algorithms, while a rate $1/2$ in Wasserstein-1 distance is obtained in [8] by using the contraction results developed in [14].

As for the generalisation of the smoothness of H, to the best of the author's knowledge, there are no theoretical guarantees established in the literature for the SGLD algorithm (2) with discontinuous gradient. We present here the first such results. We are inspired by similar studies for stochastic gradient descent (SGD) algorithms, see [15] and [7] and references therein. In particular, [15] provides an almost sure convergence result, while [7] provides a strong L_1 convergence result with rate $1/2$.

In this paper, we establish non-asymptotic error bounds for the SGLD algorithm (2) with discontinuous gradient H. More precisely, non-asymptotic results in Wasserstein-1 and Wasserstein-2 distances between the law of the n-th iterate of the SGLD algorithm (2) and the target distribution π_β are obtained under convexity and dissipativity conditions for H. This allows us to then provide full analytic results concerning the expected excess risk of the associated optimization problem (1). All this is achieved by assuming that H is decomposed in to two parts F and G, where $F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ is locally Lipschitz continuous and $G : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ is bounded. Furthermore, H is assumed to satisfy a conditional Lipschitz-continuity (CLC) property proposed in [7], which is given explicitly in Assumption 3 below.

We illustrate the applicability of our findings by presenting examples from quantile and VaR, CVaR estimations in Section 5. In particular, we solve the problem of optimal allocation of weights for the minimization of CVaR of a portfolio of assets. This is also the first such result in the literature to the best of the author's knowledge. Numerical experiments are implemented and their results support our theoretical findings.

The paper is organised as follows. Section 2 presents the assumptions and main results. In Section 3, the proofs for the main theorems in the non-convex case are provided, which are followed by the proofs for the results in the convex case in Section 4. Practical examples along with the minimization algorithm of CVaR for a portfolio of assets are presented in Section 5 while auxiliary results are provided in Section A.

We conclude this section by introducing some notation. Let $(\Omega, \mathcal{F}, P)$ be a probability space. We denote by $\mathbb{E}[X]$ the expectation of a random variable X. For any $x \in \mathbb{R}^d$, denote by $x^{(i)}$ the i-th entry of the vector. Fix an integer $d \geq 1$. For an $\mathbb{R}^d$-valued random variable X, its law on $\mathcal{B}(\mathbb{R}^d)$ (the Borel sigma-algebra of $\mathbb{R}^d$) is denoted by $\mathcal{L}(X)$. Scalar product is denoted by $\langle \cdot, \cdot \rangle$, with $|\cdot|$ standing for the corresponding norm (where the dimension of the space may vary depending on the context). For $\mu \in \mathcal{P}(\mathbb{R}^d)$ and for a non-negative measurable $f : \mathbb{R}^d \to \mathbb{R}$, the notation $\mu(f) := \int_{\mathbb{R}^d} f(\theta)\mu(d\theta)$ is used. Given a Markov kernel R on $\mathbb{R}^d$ and a function f integrable under $R(x, \cdot)$, for any $x \in \mathbb{R}^d$, denote by $Rf(x) = \int_{\mathbb{R}^d} f(y)R(x, dy)$. For any integer $q \geq 1$, let $\mathcal{P}(\mathbb{R}^q)$ denote the set of probability measures on $\mathcal{B}(\mathbb{R}^q)$. For $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, let $\mathcal{C}(\mu, \nu)$ denote the set of probability measures ζ on $\mathcal{B}(\mathbb{R}^{2d})$ such that its respective marginals are μ, ν. For two probability measures μ and ν, the Wasserstein distance of order $p \geq 1$ is defined as

$$W_p(\mu, \nu) := \inf_{\zeta \in \mathcal{C}(\mu,\nu)} \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\theta - \theta'|^p \zeta(d\theta d\theta') \right)^{1/p}, \quad \mu, \nu \in \mathcal{P}(\mathbb{R}^d). \tag{4}$$

2 Main results

Denote by $\mathcal{G}_n := \sigma(X_k,\ k \leq n,\ k \in \mathbb{N})$, for any $n \in \mathbb{N}$. $(X_n)_{n\in\mathbb{N}}$ is an $\mathbb{R}^m$-valued, $(\mathcal{G}_n)_{n\in\mathbb{N}}$-adapted process. It is assumed throughout the paper that θ_0, $\mathcal{G}_\infty$ and $(\xi_n)_{n\in\mathbb{N}}$ are independent. Moreover, the following assumptions are considered:

Assumption 1. *Let $H : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ take the form*

$$H(\theta, x) = F(\theta, x) + G(\theta, x), \quad \theta \in \mathbb{R}^d, \quad x \in \mathbb{R}^m,$$

where $F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ and $G : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ satisfy the following:

(i) $F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ is jointly Lipschitz continuous in both variables, i.e. there exist $L_1, L_2 > 0$, $\rho \geq 0$ such that for any $\theta, \theta' \in \mathbb{R}^d$, $x, x' \in \mathbb{R}^m$,

$$|F(\theta, x) - F(\theta', x')| \leq (1 + |x| + |x'|)^\rho (L_1|\theta - \theta'| + L_2|x - x'|).$$

(ii) $G(\theta, x) : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ is bounded in θ, i.e. there exist $K_1 : \mathbb{R}^m \to \mathbb{R}_+$ such that for any $\theta \in \mathbb{R}^d$, $x \in \mathbb{R}^m$,

$$|G(\theta, x)| \leq K_1(x).$$

Assumption 2. *We assume the inital value θ_0 satisfies $\mathbb{E}[|\theta_0|^4] < \infty$. The process $(X_n)_{n\in\mathbb{N}}$ is i.i.d. with $\mathbb{E}[|X_0|^{4\rho+4}] < \infty$ and $\mathbb{E}[K_1^4(X_0)] < \infty$. Moreover, it satisfies*

$$\mathbb{E}[H(\theta, X_0)] = h(\theta).$$

Remark 1. *By Assumption 1, for all $\theta \in \mathbb{R}^d$ and $x \in \mathbb{R}^m$,*

$$|H(\theta, x)| \leq (1 + |x|)^{\rho+1}(L_1|\theta| + L_2) + F_*(x),$$

where $F_*(x) = |F(0, 0)| + K_1(x)$. For any $x \in \mathbb{R}^m$, $\rho \geq 0$, denote by

$$K_\rho(x) = (1 + 2|x|)^{4\rho+4}. \tag{5}$$

One notices that by Assumption 2, $\mathbb{E}[K_\rho(X_0)]$ is well defined.

Assumption 3. *There exists a positive constant $L > 0$ such that, for all $\theta, \theta' \in \mathbb{R}^d$,*

$$\mathbb{E}[|H(\theta, X_0) - H(\theta', X_0)|] \leq L|\theta - \theta'|.$$

Remark 2. *Assumptions 2 and 3 imply, for all $\theta, \theta' \in \mathbb{R}^d$,*

$$|h(\theta) - h(\theta')| \leq L|\theta - \theta'|. \tag{6}$$

Remark 3. *Assumption 3 is satisfied for a wide class of $(X_n)_{n\in\mathbb{N}}$, see Section 5 for the examples. Here, for the illustrative purpose, one considers the following simple example. Suppose $G(\theta, x) = \sum_{j=1}^N \dot{g}_j(\theta, x) \mathbb{1}_{\bigcap_{i=1}^m \{x^{(i)} \in I_{i,j}(\theta)\}}$ is a lower semi-continuous function, where $N \in \mathbb{N}^*$, $\dot{g}_j : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ are bounded and jointly Lipschitz continuous functions, i.e. there exist $L_3, L_4, K_2 > 0$ such that for any $\theta, \theta' \in \mathbb{R}^d$, $x, x' \in \mathbb{R}^m$, $j = 1, \ldots, N$*

$$|\dot{g}_j(\theta, x) - \dot{g}_j(\theta', x')| \leq (1 + |x| + |x'|)^\rho (L_3|\theta - \theta'| + L_4|x - x'|), \quad |\dot{g}_j(\theta, x)| \leq K_2,$$

the intervals $I_{i,j}(\theta)$ take the form $(-\infty, \bar{g}_j^{(i)}(\theta))$, $(\bar{g}_j^{(i)}(\theta), \infty)$ or $(\tilde{g}_j^{(i)}(\theta), \hat{g}_j^{(i)}(\theta))$, and $\bar{g}_j^{(i)}, \tilde{g}_j^{(i)}, \hat{g}_j^{(i)} : \mathbb{R}^d \to \mathbb{R}$ are Lipschitz continuous functions. In this case, it is enough to require the marginal density function of $X_0^{(i)}$ is continuous and bounded for any $i = 1, \ldots, m$. Then, the property stated in Assumption 3 holds.

Proof. See Appendix A.1. □

2.1 Nonconvex case

Further to the assumptions above, we consider the following conditions on U, which can be viewed as a generalization of the convexity assumption.

Assumption 4. *There exist $A : \mathbb{R}^m \to \mathbb{R}^{d\times d}$, $b : \mathbb{R}^m \to \mathbb{R}$ such that for any $x, y \in \mathbb{R}^d$,*

$$\langle y, A(x)y\rangle \geq 0$$

and for all $\theta \in \mathbb{R}^d$ and $x \in \mathbb{R}^m$,

$$\langle F(\theta, x), \theta\rangle \geq \langle \theta, A(x)\theta\rangle - b(x).$$

The smallest eigenvalue of $\mathbb{E}[A(X_0)]$ is a positive real number $a > 0$ and $E[b(X_0)] = b > 0$.

Define first

$$\lambda_{\max} = \min\left\{\frac{\min\{a, a^{1/3}\}}{24(1+L_1)^2\mathbb{E}\left[K_\rho(X_0)\right]}, \frac{1}{4a}\right\}, \tag{7}$$

where L_1, a are given in Assumption 1 and 4 respectively, and $K_\rho(x)$ for any $x \in \mathbb{R}^m$ is defined in (5).

Theorem 1. *Let Assumptions 1, 2, 3 and 4 hold. Then, for any $n \in \mathbb{N}$, $0 < \lambda \leq \lambda_{\max}$, there exist constants $C_0, C_1, C_2 > 0$ such that,*

$$W_1(\mathcal{L}(\theta_n^\lambda), \pi_\beta) \leq C_1 e^{-C_0\lambda n}(\mathbb{E}[|\theta_0|^4] + 1) + C_2\sqrt{\lambda},\ n \in \mathbb{N}, \tag{8}$$

where C_0, C_1 and C_2 are given explicitly in (29)*.*

Theorem 1 provides the rate of convergence between the law of the SGLD algorithm (2) and the target distribution π_β in W_1 distance. An analogous result in Wasserstein-2 distance can be obtained.

Corollary 1. *Let Assumptions 1, 2, 3 and 4 hold. Then, for any $n \in \mathbb{N}$, $0 < \lambda \leq \lambda_{\max}$ given in* (7)*, there exist constants $C_3, C_4, C_5 > 0$ such that,*

$$W_2(\mathcal{L}(\theta_n^\lambda), \pi_\beta) \leq C_4 e^{-C_3\lambda n}(\mathbb{E}[|\theta_0|^4] + 1) + C_5\lambda^{1/4},\ n \in \mathbb{N},$$

where C_3, C_4 and C_5 are given explicitly in (30)*.*

By using the convergence result in Wasserstein-2 distance as presented in Corollary 1, one can obtain an upper bound for the expected excess risk $\mathbb{E}[U(\hat{\theta})] - \inf_{\theta\in\mathbb{R}^d} U(\theta)$.

Corollary 2. *Let Assumptions 1, 2, 3 and 4 hold. Then, for every $0 < \lambda \leq \lambda_{\max}$ given in* (7)*, there exist constants $\hat{C}_0, \hat{C}_1, \hat{C}_2, \hat{C}_3 > 0$ such that the expected excess risk can be estimated as*

$$\mathbb{E}[U(\hat{\theta})] - \inf_{\theta\in\mathbb{R}^d} U(\theta) \leq \hat{C}_1 e^{-\hat{C}_0\lambda n} + \hat{C}_2\lambda^{1/4} + \hat{C}_3/\beta,$$

where $\hat{\theta} = \theta_n^\lambda$, and $\hat{C}_0, \hat{C}_1, \hat{C}_2$, $\hat{C}_3 > 0$ are given explicitly in (32) *and* (33)*.*

2.2 Convex case

Recall Assumption 1, where it is assumed $H = F + G$. In this section, we present (improved) convergence results of the SGLD algorithm (2) under the convexity condition of F and G.

In the case that F satisfies a convexity condition but not G, the result in Theorem 1 can be recovered.

Assumption 5. *There exist $\hat{A}_1 : \mathbb{R}^m \to \mathbb{R}^{d\times d}$ such that for any $x, y \in \mathbb{R}^d$,*

$$\langle y, \hat{A}_1(x)y\rangle \geq 0$$

and for each $\theta, \theta' \in \mathbb{R}^d$, $x \in \mathbb{R}^m$,

$$\langle F(\theta, x) - F(\theta', x), \theta - \theta'\rangle \geq \langle \theta - \theta', \hat{A}_1(x)(\theta - \theta')\rangle.$$

The smallest eigenvalue of $\mathbb{E}[\hat{A}_1(X_0)]$ is a positive real number $\hat{a}_1 > \epsilon$ with $\epsilon > 0$.

Remark 4. *By Assumptions 1 and 5, one obtains, for $\theta \in \mathbb{R}^d$ and $x \in \mathbb{R}^m$,*

$$\langle F(\theta, x), \theta \rangle \geq \langle \theta, \hat{A}_1^*(x)\theta \rangle - \hat{b}(x),$$

where $\hat{A}_1^*(x) = \hat{A}_1(x) - \epsilon \mathbf{I}_d$ and $\hat{b}(x) = (L_2(1+|x|)^{\rho+1} + |F(0,0)|)^2/(4\epsilon)$.

Proof. See Appendix A.2. □

Corollary 3. *Let Assumptions 1, 2, 3 and 5 hold. Then, for any $n \in \mathbb{N}$, $0 < \lambda \leq \lambda_{\max}^*$, where*

$$\lambda_{\max}^* = \min\left\{ \frac{\min\{a^*, (a^*)^{1/3}\}}{24(1+L_1)^2 \mathbb{E}\left[K_\rho(X_0)\right]}, \frac{1}{4a^*} \right\}$$

with $a^* = \hat{a}_1 - \epsilon$, there exist constants $C_0^*, C_1^*, C_2^* > 0$ such that,

$$W_1(\mathcal{L}(\theta_n^\lambda), \pi_\beta) \leq C_1^* e^{-C_0^* \lambda n}(\mathbb{E}[|\theta_0|^4] + 1) + C_2^* \sqrt{\lambda}, \ n \in \mathbb{N}. \tag{9}$$

If G is assumed to be convex in addition to Assumption 5, then it can be shown that the rate of convergence is 1/2 in Wasserstein-2 distance between the law of the SGLD algorithm (2) and the target distribution π_β, which appeared to be optimal, see [1, Example 3.4].

Assumption 6. *There exist $\hat{A}_2 : \mathbb{R}^m \to \mathbb{R}^{d \times d}$ such that for any $x, y \in \mathbb{R}^d$,*

$$\langle y, \hat{A}_2(x) y \rangle \geq 0$$

and for each $\theta, \theta' \in \mathbb{R}^d$, $x \in \mathbb{R}^m$,

$$\langle G(\theta, x) - G(\theta', x), \theta - \theta' \rangle \geq \langle \theta - \theta', \hat{A}_2(x)(\theta - \theta') \rangle.$$

The smallest eigenvalue of $\mathbb{E}[\hat{A}_2(X_0)]$ is a positive real number $\hat{a}_2 > 0$.

Remark 5. *Assumptions 5 and 6 imply, for each $\theta, \theta' \in \mathbb{R}^d$, $x \in \mathbb{R}^m$,*

$$\langle H(\theta, x) - H(\theta', x), \theta - \theta' \rangle \geq \langle \theta - \theta', \hat{A}(x)(\theta - \theta') \rangle,$$

where $\hat{A}(x) = \hat{A}_1(x) + \hat{A}_2(x)$. Moreover, one obtains

$$\langle h(\theta) - h(\theta'), \theta - \theta' \rangle \geq \hat{a}|\theta - \theta'|^2,$$

where $\hat{a} = \hat{a}_1 + \hat{a}_2$.

Remark 6. *By Remark 2 and Remark 5, [18, Theorem 2.1.12] shows that*

$$\langle h(\theta) - h(\theta'), \theta - \theta' \rangle \geq \hat{a}^* |\theta - \theta'|^2 + \frac{1}{\hat{a} + L}|h(\theta) - h(\theta')|^2,$$

where $\hat{a}^* = \hat{a}L/(\hat{a} + L)$.

Define

$$\bar{\lambda}_{\max} = \min\{1/2(\hat{a} + L), \hat{a}/(4L_1^2 \mathbb{E}[K_\rho(X_0)])\} \tag{10}$$

with $\hat{a} = \hat{a}_1 + \hat{a}_2$ given in Remark 5. Under the convexity condition of H, the non-asymptotic bound for $W_2(\mathcal{L}(\theta_n^\gamma), \pi_\beta)$ is obtained with the optimal convergence rate 1/2. The explicit statement is given below.

Theorem 2. *Let Assumptions 1, 2, 3, 5 and 6 hold. Then, for any $n \in \mathbb{N}$, $0 < \lambda < \bar{\lambda}_{\max}$ given in* (10)*, there exist constants $C_6, C_7, C_8 > 0$ such that,*

$$W_2(\mathcal{L}(\theta_n^\lambda), \pi_\beta) \leq C_7 e^{-C_6 \lambda n} + C_8 \sqrt{\lambda},$$

where C_6, C_7 and C_8 are given explicitly in (41)*. If $\rho = 0$ in Assumption 1, then the result holds for $\lambda \in \min\{1/2(\hat{a} + L), 1/(6L_1)\}$.*

By using Theorem 2, one can obtain an upper bound for the expected excess risk $\mathbb{E}[U(\hat{\theta})] - \inf_{\theta \in \mathbb{R}^d} U(\theta)$ in the convex case.

Corollary 4. *Let Assumptions 1, 2, 3, 5 and 6 hold. Then, for every $0 < \lambda \leq \bar{\lambda}_{\max}$ given in* (10)*, there exist constants $\hat{C}_4, \hat{C}_5, \hat{C}_6, \hat{C}_7 > 0$ such that the expected excess risk can be estimated as*

$$\mathbb{E}[U(\hat{\theta})] - \inf_{\theta \in \mathbb{R}^d} U(\theta) \leq \hat{C}_5 e^{-\hat{C}_4 \lambda n} + \hat{C}_6 \sqrt{\lambda} + \hat{C}_7/\beta,$$

where $\hat{\theta} = \theta_n^\lambda$, and $\hat{C}_4, \hat{C}_5, \hat{C}_6, \hat{C}_7 > 0$ are given explicitly in (43) *and* (44).

3 Proofs of the main results: nonconvex case

Denote by $\mathcal{F}_t$ the natural filtration of B_t, $t \in \mathbb{R}_+$. It is a classic result that SDE (3) has a unique solution adapted to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, since h is Lipschitz-continuous by (6). In order to obtain the convergence results in Theorem 1 and Corollary 1, we first introduce some auxiliary processes.

3.1 Further notation and introduction of auxiliary processes

Define the Lyapunov function for each $p \geq 1$ by

$$V_p(\theta) := (1+|\theta|^2)^{p/2},\ \theta \in \mathbb{R}^d,$$

and similarly $\mathrm{v}_p(x) := (1+x^2)^{p/2}$, for any real $x \geq 0$. Notice that these functions are twice continuously differentiable and

$$\lim_{|\theta|\to\infty} \frac{\nabla V_p(\theta)}{V_p(\theta)} = 0.$$

Let $\mathcal{P}_{V_p}$ denote the set of $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfying $\int_{\mathbb{R}^d} V_p(\theta)\,\mu(d\theta) < \infty$.

Consider the following auxiliary processes. For each $\lambda > 0$,

$$Z_t^\lambda := Z_{\lambda t},\ t \in \mathbb{R}_+.$$

Notice that $\tilde{B}_t^\lambda := B_{\lambda t}/\sqrt{\lambda}$, $t \in \mathbb{R}_+$ is also a Brownian motion and

$$dZ_t^\lambda = -\lambda h(Z_t^\lambda)\,dt + \sqrt{2\beta^{-1}\lambda}d\tilde{B}_t^\lambda,\ Z_0^\lambda = \theta_0.$$

Then, $\mathcal{F}_t^\lambda := \mathcal{F}_{\lambda t}$, $t \in \mathbb{R}_+$ is the natural filtration of $\tilde{B}_t^\lambda$, $t \in \mathbb{R}_+$. One notice that $\mathcal{F}_t^\lambda$ is independent of $\mathcal{G}_\infty \vee \sigma(\theta_0)$. Then, define the continuous-time interpolation of the SGLD algorithm (2) as

$$d\bar{\theta}_t^\lambda = -\lambda H(\bar{\theta}_{\lfloor t\rfloor}^\lambda, X_{\lceil t\rceil})\,dt + \sqrt{2\beta^{-1}\lambda}d\tilde{B}_t^\lambda, \tag{11}$$

with initial condition $\bar{\theta}_0^\lambda = \theta_0$. In addition, due to the homogeneous nature of the coefficients of equation (11), the law of the interpolated process coincides with the law of the SGLD algorithm (2) at grid-points, i.e. $\mathcal{L}(\bar{\theta}_n^\lambda) = \mathcal{L}(\theta_n^\lambda)$, for each $n \in \mathbb{N}$. Hence, crucial estimates for the SGLD can be derived by studying equation (11).

Furthermore, consider a continuous-time process $\zeta_t^{s,v,\lambda}$, $t \geq s$, which denotes the solution of the SDE

$$d\zeta_t^{s,v,\lambda} = -\lambda h(\zeta_t^{s,v,\lambda})dt + \sqrt{2\beta^{-1}\lambda}d\tilde{B}_t^\lambda.$$

with initial condition $\zeta_s^{s,v,\lambda} := v$, $v \in \mathbb{R}^d$.

Definition 1. *Fix $n \in \mathbb{N}$ and define*

$$\bar{\zeta}_t^{\lambda,n} = \zeta_t^{nT,\bar{\theta}_{nT}^\lambda,\lambda}$$

where $T := \lfloor 1/\lambda \rfloor$.

Intuitively, $\bar{\zeta}_t^{\lambda,n}$ is a process started from the value of the SGLD process (11) at time nT and made run until time $t \geq nT$ with the continuous-time Langevin dynamics.

3.2 Preliminary estimates

We proceed by establishing the moment bounds of the processes $(\bar{\theta}_t^\lambda)_{t\geq 0}$ and $(\bar{\zeta}_t^{\lambda,n})_{t\geq 0}$.

Lemma 1. *Let Assumptions 1, 2 and 4 hold. For any $0 < \lambda < \lambda_{\max}$ given in (7), $n \in \mathbb{N}$, $t \in (n, n+1]$,*

$$\mathbb{E}\left[|\bar{\theta}_t^\lambda|^2\right] \leq (1-a\lambda(t-n))(1-a\lambda)^n \mathbb{E}\left[|\theta_0|^2\right] + c_1(\lambda_{\max} + a^{-1}),$$

where

$$c_1 = (c_0 + 2d/\beta), \quad c_0 = 8\mathbb{E}\left[K_1^2(X_0)\right] a^{-1} + 2b + 4\lambda_{\max} L_2^2 \mathbb{E}\left[K_\rho(X_0)\right] + 4\lambda_{\max}\mathbb{E}\left[F_*^2(X_0)\right]. \tag{12}$$

In addition, $\sup_t \mathbb{E}|\bar{\theta}_t^\lambda|^2 \leq \mathbb{E}\left[|\theta_0|^2\right] + c_1(\lambda_{\max} + a^{-1}) < \infty$. *Similarly, one obtains*

$$\mathbb{E}\left[|\bar{\theta}_t^\lambda|^4\right] \leq (1 - a\lambda(t-n))(1-a\lambda)^n \mathbb{E}|\theta_0|^4 + c_3(\lambda_{\max} + a^{-1}),$$

where

$$c_3 = (1 + a\lambda_{\max})c_2 + 12d^2\beta^{-2}(\lambda_{\max} + 9a^{-1}) \tag{13}$$

with c_2 *given in* (18)*. Moreover, this implies* $\sup_t \mathbb{E}|\bar{\theta}_t^\lambda|^4 < \infty$.

Proof. For any $n \in \mathbb{N}$ and $t \in (n, n+1]$, define $\Delta_{n,t} = \bar{\theta}_n^\lambda - \lambda H(\bar{\theta}_n^\lambda, X_{n+1})(t-n)$. By using (11), it is easily seen that for $t \in (n, n+1]$

$$\mathbb{E}\left[|\bar{\theta}_t^\lambda|^2 \middle| \bar{\theta}_n^\lambda\right] = \mathbb{E}\left[|\Delta_{n,t}|^2 \middle| \bar{\theta}_n^\lambda\right] + (2\lambda/\beta)d(t-n).$$

Then, by using Assumptions 1, 2, 4 and Remark 1, one obtains

$$\begin{aligned}
\mathbb{E}\left[|\Delta_{n,t}|^2 \middle| \bar{\theta}_n^\lambda\right] &= |\bar{\theta}_n^\lambda|^2 - 2\lambda(t-n)\mathbb{E}\left[\left\langle \bar{\theta}_n^\lambda, H(\bar{\theta}_n^\lambda, X_{n+1})\right\rangle \middle| \bar{\theta}_n^\lambda\right] \\
&\quad + \lambda^2(t-n)^2 \mathbb{E}\left[|H(\bar{\theta}_n^\lambda, X_{n+1})|^2 \middle| \bar{\theta}_n^\lambda\right] \\
&\leq |\bar{\theta}_n^\lambda|^2 - 2\lambda(t-n)\left\langle \bar{\theta}_n^\lambda, \mathbb{E}\left[A(X_0)\right]\bar{\theta}_n^\lambda\right\rangle + 2\lambda(t-n)b \\
&\quad - 2\lambda(t-n)\mathbb{E}\left[\left\langle \bar{\theta}_n^\lambda, G(\bar{\theta}_n^\lambda, X_{n+1})\right\rangle \middle| \bar{\theta}_n^\lambda\right] \\
&\quad + \lambda^2(t-n)^2\mathbb{E}\left[((1+|X_{n+1}|)^{\rho+1}(L_1|\bar{\theta}_n^\lambda| + L_2) + F_*(X_{n+1}))^2 \middle| \bar{\theta}_n^\lambda\right] \\
&\leq (1 - 2a\lambda(t-n))|\bar{\theta}_n^\lambda|^2 + 2\lambda(t-n)b + 2\lambda(t-n)\mathbb{E}\left[K_1(X_0)\right]|\bar{\theta}_n^\lambda| \\
&\quad + 2\lambda^2(t-n)^2 L_1^2 \mathbb{E}\left[K_\rho(X_0)\right]|\bar{\theta}_n^\lambda|^2 + 4\lambda^2(t-n)^2 L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + 4\lambda^2(t-n)^2\mathbb{E}\left[F_*^2(X_0)\right],
\end{aligned}$$

where the last inequality is obtained by using $(a+b)^2 \leq 2a^2 + 2b^2$, for $a, b \geq 0$ twice. For $\lambda < \lambda_{\max}$ with $\lambda_{\max}$ given in (7),

$$\begin{aligned}
\mathbb{E}\left[|\Delta_{n,t}|^2 \middle| \bar{\theta}_n^\lambda\right] &\leq \left(1 - \frac{3}{2}a\lambda(t-n)\right)|\bar{\theta}_n^\lambda|^2 + 2\lambda(t-n)\mathbb{E}\left[K_1(X_0)\right]|\bar{\theta}_n^\lambda| \\
&\quad + 2\lambda(t-n)b + 4\lambda^2(t-n)^2 L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + 4\lambda^2(t-n)^2\mathbb{E}[F_*^2(X_0)].
\end{aligned}$$

For $|\bar{\theta}_n^\lambda| > 4\mathbb{E}\left[K_1(X_0)\right]a^{-1}$, one obtains

$$-\frac{1}{2}a\lambda(t-n)|\bar{\theta}_n^\lambda|^2 + 2\lambda(t-n)\mathbb{E}\left[K_1(X_0)\right]|\bar{\theta}_n^\lambda| < 0,$$

which implies

$$\begin{aligned}
\mathbb{E}\left[|\Delta_{n,t}|^2 \middle| \bar{\theta}_n^\lambda\right] &\leq (1 - a\lambda(t-n))\,|\bar{\theta}_n^\lambda|^2 + 2\lambda(t-n)b \\
&\quad + 4\lambda^2(t-n)^2 L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + 4\lambda^2(t-n)^2\mathbb{E}\left[F_*^2(X_0)\right].
\end{aligned}$$

For $|\bar{\theta}_n^\lambda| \leq 4\mathbb{E}\left[K_1(X_0)\right]a^{-1}$, we have

$$\begin{aligned}
\mathbb{E}\left[|\Delta_{n,t}|^2 \middle| \bar{\theta}_n^\lambda\right] &\leq \left(1 - \frac{3}{2}a\lambda(t-n)\right)|\bar{\theta}_n^\lambda|^2 + 8\lambda(t-n)\mathbb{E}\left[K_1^2(X_0)\right]a^{-1} \\
&\quad + 2\lambda(t-n)b + 4\lambda^2(t-n)^2 L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + 4\lambda^2(t-n)^2\mathbb{E}\left[F_*^2(X_0)\right].
\end{aligned}$$

Combining the two cases yields

$$\mathbb{E}\left[|\Delta_{n,t}|^2 \middle| \bar{\theta}_n^\lambda\right] \leq (1 - a\lambda(t-n))\,|\bar{\theta}_n^\lambda|^2 + \lambda(t-n)c_0,$$

where $c_0 = 8\mathbb{E}\left[K_1^2(X_0)\right] a^{-1} + 2b + 4\lambda_{\max} L_2^2 \mathbb{E}\left[K_\rho(X_0)\right] + 4\lambda_{\max}\mathbb{E}\left[F_*^2(X_0)\right]$. Therefore, one obtains

$$\mathbb{E}\left[|\bar{\theta}_t^\lambda|^2 \middle| \bar{\theta}_n^\lambda\right] \leq (1 - a\lambda(t-n))|\bar{\theta}_n^\lambda|^2 + \lambda(t-n)c_1,$$

where $c_1 = (c_0 + 2d/\beta)$ and the result follows by induction. To calculate a higher moment, denote by $\Xi_{n,t}^\lambda = \{2\lambda\beta^{-1}\}^{1/2}(\tilde{B}_t^\lambda - \tilde{B}_n^\lambda)$, for $t \in (n, n+1]$, one calculates

$$\begin{aligned}
\mathbb{E}\left[|\bar{\theta}_t^\lambda|^4 \middle| \bar{\theta}_n^\lambda\right] &= \mathbb{E}\left[\left(|\Delta_{n,t}|^2 + |\Xi_{n,t}^\lambda|^2 + 2\left\langle \Delta_{n,t}, \Xi_{n,t}^\lambda\right\rangle\right)^2 \middle| \bar{\theta}_n^\lambda\right] \\
&= \mathbb{E}\left[|\Delta_{n,t}|^4 + |\Xi_{n,t}^\lambda|^4 + 2|\Delta_{n,t}|^2|\Xi_{n,t}^\lambda|^2 + 4|\Delta_{n,t}|^2\left\langle \Delta_{n,t}, \Xi_{n,t}^\lambda\right\rangle\right. \\
&\qquad \left. +4|\Xi_{n,t}^\lambda|^2\left\langle \Delta_{n,t}, \Xi_{n,t}^\lambda\right\rangle + 4\left(\left\langle \Delta_{n,t}, \Xi_{n,t}^\lambda\right\rangle\right)^2 \middle| \bar{\theta}_n^\lambda\right] \\
&\leq \mathbb{E}\left[|\Delta_{n,t}|^4 + |\Xi_{n,t}^\lambda|^4 + 6|\Delta_{n,t}|^2|\Xi_{n,t}^\lambda|^2 \middle| \bar{\theta}_n^\lambda\right] \\
&\leq (1 + a\lambda(t-n))\mathbb{E}\left[|\Delta_{n,t}|^4 \middle| \bar{\theta}_n^\lambda\right] + (1 + 9/(a\lambda(t-n)))\mathbb{E}\left[|\Xi_{n,t}^\lambda|^4\right].
\end{aligned} \tag{14}$$

where the last inequality holds due to $2ab \leq \varepsilon a^2 + \varepsilon^{-1}b^2$, for $a, b \geq 0$ and $\varepsilon > 0$ with $\varepsilon = a\lambda(t-n)$. Then, one continues with calculating

$$\begin{aligned}
\mathbb{E}\left[|\Delta_{n,t}|^4 \middle| \bar{\theta}_n^\lambda\right] &= \mathbb{E}\left[\left(|\bar{\theta}_n^\lambda|^2 - 2\lambda(t-n)\left\langle \bar{\theta}_n^\lambda, H(\bar{\theta}_n^\lambda, X_{n+1})\right\rangle + \lambda^2(t-n)^2|H(\bar{\theta}_n^\lambda, X_{n+1})|^2\right)^2 \middle| \bar{\theta}_n^\lambda\right] \\
&\leq |\bar{\theta}_n^\lambda|^4 + \mathbb{E}\left[6\lambda^2(t-n)^2|\bar{\theta}_n^\lambda|^2|H(\bar{\theta}_n^\lambda, X_{n+1})|^2 - 4\lambda(t-n)\left\langle \bar{\theta}_n^\lambda, H(\bar{\theta}_n^\lambda, X_{n+1})\right\rangle|\bar{\theta}_n^\lambda|^2\right. \\
&\qquad \left. -4\lambda^3(t-n)^3|H(\bar{\theta}_n^\lambda, X_{n+1})|^2\left\langle \bar{\theta}_n^\lambda, H(\bar{\theta}_n^\lambda, X_{n+1})\right\rangle + \lambda^4(t-n)^4|H(\bar{\theta}_n^\lambda, X_{n+1})|^4 \middle| \bar{\theta}_n^\lambda\right].
\end{aligned}$$

By Remark 1, for $q \geq 1$, one observes

$$\mathbb{E}\left[|H(\bar{\theta}_n^\lambda, X_{n+1})|^q \middle| \bar{\theta}_n^\lambda\right] \leq \mathbb{E}\left[(1 + |X_0|)^{q\rho+q}\right](2^{q-1}L_1^q|\bar{\theta}_n^\lambda|^q + 2^{2q-2}L_2^q) + 2^{2q-2}\mathbb{E}\left[F_*^q(X_0)\right]. \tag{15}$$

Then, by using Assumption 4 and by taking $q = 2, 3, 4$ in (15), one obtains

$$\begin{aligned}
&\mathbb{E}\left[|\Delta_{n,t}|^4 \middle| \bar{\theta}_n^\lambda\right] \\
&\leq (1 - 4a\lambda(t-n))|\bar{\theta}_n^\lambda|^4 + 4b\lambda(t-n)|\bar{\theta}_n^\lambda|^2 + 4\lambda(t-n)\mathbb{E}\left[K_1(X_0)\right]|\bar{\theta}_n^\lambda|^3 \\
&\quad + 12\lambda^2(t-n)^2L_1^2\mathbb{E}\left[K_\rho(X_0)\right]|\bar{\theta}_n^\lambda|^4 + 24\lambda^2(t-n)^2\left(L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^2(X_0)\right]\right)|\bar{\theta}_n^\lambda|^2 \\
&\quad + 16\lambda^3(t-n)^3L_1^3\mathbb{E}\left[K_\rho(X_0)\right]|\bar{\theta}_n^\lambda|^4 + 64\lambda^3(t-n)^3\left(L_2^3\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^3(X_0)\right]\right)|\bar{\theta}_n^\lambda| \\
&\quad + 8\lambda^4(t-n)^4L_1^4\mathbb{E}\left[K_\rho(X_0)\right]|\bar{\theta}_n^\lambda|^4 + 64\lambda^4(t-n)^4\left(L_2^4\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^4(X_0)\right]\right),
\end{aligned}$$

which implies, by using $\lambda < \lambda_{\max}$

$$\begin{aligned}
\mathbb{E}\left[|\Delta_{n,t}|^4 \middle| \bar{\theta}_n^\lambda\right] &\leq (1 - 3a\lambda(t-n))|\bar{\theta}_n^\lambda|^4 + 4\lambda(t-n)\mathbb{E}\left[K_1(X_0)\right]|\bar{\theta}_n^\lambda|^3 \\
&\quad + 4b\lambda(t-n)|\bar{\theta}_n^\lambda|^2 + 24\lambda^2(t-n)^2\left(L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^2(X_0)\right]\right)|\bar{\theta}_n^\lambda|^2 \\
&\quad + 64\lambda^3(t-n)^3\left(L_2^3\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^3(X_0)\right]\right)|\bar{\theta}_n^\lambda| \\
&\quad + 64\lambda^4(t-n)^4\left(L_2^4\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^4(X_0)\right]\right).
\end{aligned}$$

For $|\bar{\theta}_n^\lambda| > 12\mathbb{E}\left[K_1(X_0)\right]a^{-1}$, one obtains

$$-\frac{a\lambda(t-n)}{3}|\bar{\theta}_n^\lambda|^4 + 4\lambda(t-n)\mathbb{E}\left[K_1(X_0)\right]|\bar{\theta}_n^\lambda|^3 < 0,$$

similarly, for $|\bar{\theta}_n^\lambda| > (12ba^{-1} + 72a^{-1}\lambda_{\max}\left(L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^2(X_0)\right]\right))^{1/2}$, we have

$$-\frac{a\lambda(t-n)}{3}|\bar{\theta}_n^\lambda|^4 + 4b\lambda(t-n)|\bar{\theta}_n^\lambda|^2 + 24\lambda^2(t-n)^2\left(L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^2(X_0)\right]\right)|\bar{\theta}_n^\lambda|^2 < 0,$$

moreover, for $|\bar{\theta}_n^\lambda| > (192a^{-1}\lambda_{\max}^2 \left(L_2^3\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^3(X_0)\right]\right))^{1/3}$

$$-\frac{a\lambda(t-n)}{3}|\bar{\theta}_n^\lambda|^4 + 64\lambda^3(t-n)^3\left(L_2^3\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^3(X_0)\right]\right)|\bar{\theta}_n^\lambda| < 0.$$

Denote by

$$\begin{aligned} M = \max\Big\{&12\mathbb{E}\left[K_1(X_0)\right]a^{-1}, (12ba^{-1} + 72a^{-1}\lambda_{\max}\left(L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^2(X_0)\right]\right))^{1/2}, \\ &(192a^{-1}\lambda_{\max}^2\left(L_2^3\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^3(X_0)\right]\right))^{1/3}\Big\}. \end{aligned} \tag{16}$$

For $|\bar{\theta}_n^\lambda| > M$, one obtains

$$\mathbb{E}\left[|\Delta_{n,t}|^4 \middle| \bar{\theta}_n^\lambda\right] \leq (1 - 2a\lambda(t-n))|\bar{\theta}_n^\lambda|^4 + 64\lambda^4(t-n)^4\left(L_2^4\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^4(X_0)\right]\right).$$

As for $|\bar{\theta}_n^\lambda| \leq M$, we have

$$\begin{aligned} \mathbb{E}\left[|\Delta_{n,t}|^4 \middle| \bar{\theta}_n^\lambda\right] \leq\ & (1 - 3a\lambda(t-n))|\bar{\theta}_n^\lambda|^4 + 4\lambda(t-n)\mathbb{E}\left[K_1(X_0)\right]M^3 + 4b\lambda(t-n)M^2 \\ & + 24\lambda^2(t-n)^2\left(L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^2(X_0)\right]\right)M^2 \\ & + 64\lambda^3(t-n)^3\left(L_2^3\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^3(X_0)\right]\right)M \\ & + 64\lambda^4(t-n)^4\left(L_2^4\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^4(X_0)\right]\right). \end{aligned}$$

Combining the two cases yields

$$\mathbb{E}\left[|\Delta_{n,t}|^4 \middle| \bar{\theta}_n^\lambda\right] \leq (1 - 2a\lambda(t-n))|\bar{\theta}_n^\lambda|^4 + \lambda(t-n)c_2, \tag{17}$$

where

$$c_2 = 4\mathbb{E}\left[K_1(X_0)\right]M^3 + 4bM^2 + 152(1+\lambda_{\max})^3\left((1+L_2)^4\mathbb{E}\left[K_\rho(X_0)\right] + (1 + \mathbb{E}\left[F_*^4(X_0)\right]\right)(1+M)^2 \tag{18}$$

with M given in (16). Substituting (17) into (14), one obtains

$$\begin{aligned} \mathbb{E}\left[|\bar{\theta}_t^\lambda|^4 \middle| \bar{\theta}_n^\lambda\right] &\leq (1 + a\lambda(t-n))(1 - 2a\lambda(t-n))|\bar{\theta}_n^\lambda|^4 \\ &\quad + (1 + a\lambda(t-n))\lambda(t-n)c_2 + 12d^2\lambda^2\beta^{-2}(t-n)^2(1 + 9/(a\lambda(t-n))) \\ &\leq (1 - a\lambda(t-n))|\bar{\theta}_n^\lambda|^4 + \lambda(t-n)c_3, \end{aligned}$$

where $c_3 = (1 + a\lambda_{\max})c_2 + 12d^2\beta^{-2}(\lambda_{\max} + 9a^{-1})$. The proof completes by induction. □

Remark 7. *One notices that in Lemma 1, the step-size restriction is the following:*

$$\hat{\lambda}_{\max} = \min\left\{\frac{a}{24L_1^2\mathbb{E}\left[K_\rho(X_0)\right]}, \frac{a^{1/2}}{8(L_1^3\mathbb{E}\left[K_\rho(X_0)\right])^{1/2}}, \frac{a^{1/3}}{(32L_1^4\mathbb{E}\left[K_\rho(X_0)\right])^{1/3}}, \frac{1}{4a}\right\}.$$

Theorem 1 and Corollary 1 still hold by using $\hat{\lambda}_{\max}$. However, in order to make notation compact, the restriction is chosen to be $\lambda_{\max}$ given in (7), which can be deduced from the above expression.

Corollary 5. *Let Assumptions 1, 2 and 4 hold. For any $0 < \lambda < \lambda_{\max}$ given in (7), $n \in \mathbb{N}$, $t \in (n, n+1]$,*

$$\mathbb{E}[V_4(\bar{\theta}_t^\lambda)] \leq 2(1 - a\lambda)^{\lfloor t \rfloor}\mathbb{E}[V_4(\theta_0)] + 2c_3(\lambda_{\max} + a^{-1}) + 2,$$

where c_3 is given in (13).

Next, we present a drift condition associated with the SDE (3), which will be used to obtain the moment bounds of the process $(\bar{\zeta}_t^{\lambda,n})_{t\geq 0}$.

Lemma 2. *Let Assumptions 1, 2 and 4 hold. Then, for each $p \geq 2$, $\theta \in \mathbb{R}^d$,*

$$\frac{\Delta V_p}{\beta} - \langle h(\theta), \nabla V_p(\theta) \rangle \leq -\bar{c}(p) V_p(\theta) + \tilde{c}(p),$$

where $\bar{c}(p) = ap/4$ and $\tilde{c}(p) = (3/4)ap\mathrm{v}_{p+1}(\overline{M}_p)$ with $\overline{M}_p$ given in (19).

Proof. One notices that, by Assumptions 1 and 2, for any $\theta \in \mathbb{R}^d$, $h(\theta) = \mathbb{E}[H(\theta, X_0)] = \mathbb{E}[F(\theta, X_0) + G(\theta, X_0)]$. Then, one calculates,

$$\begin{aligned}
&\frac{\Delta V_p}{\beta} - \langle h(\theta), \nabla V_p(\theta) \rangle \\
&= \beta^{-1} p(p-2)|\theta|^2 V_{p-4}(\theta) + \beta^{-1} p d V_{p-2}(\theta) \\
&\quad - pV_{p-2}(\theta) \langle \mathbb{E}[F(\theta, X_0) + G(\theta, X_0)], \theta \rangle \\
&\leq -apV_p(\theta) + (ap + bp + \beta^{-1}p(p-2) + \beta^{-1}pd)V_{p-2}(\theta) + p\mathbb{E}\left[K_1(X_0)\right]|\theta| V_{p-2}(\theta),
\end{aligned}$$

where the last inequality is obtained due to Assumption 4. By observing $|\theta| \leq \sqrt{1+|\theta|^2}$, denote by

$$\overline{M}_p = \sqrt{(4/3 + 4b/(3a) + 4d/(3a\beta) + 4(p-2)/(3a\beta) + 4\mathbb{E}\left[K_1(X_0)\right]/(3a))^2 - 1}. \tag{19}$$

For $|\theta| > \overline{M}_p$, one obtains $\frac{\Delta V_p}{\beta} - \langle h(\theta), \nabla V_p(\theta) \rangle \leq -(ap/4)V_p(\theta)$, while for $|\theta| \leq \overline{M}_p$, we have $\frac{\Delta V_p}{\beta} - \langle h(\theta), \nabla V_p(\theta) \rangle \leq (3/4)ap\mathrm{v}_{p+1}(\overline{M}_p)$. Combining the two cases yields the desired result. $\square$

The following Lemma provides the second and the fourth moment of the process $(\bar{\zeta}_t^{\lambda,n})_{t\geq 0}$.

Lemma 3. *Let Assumptions 1, 2 and 4 hold. For any $0 < \lambda < \lambda_{\max}$ given in* (7), *$t \geq nT$, $n \in \mathbb{N}$, one obtains the following inequality*

$$\mathbb{E}[V_2(\bar{\zeta}_t^{\lambda,n})] \leq e^{-a\lambda t/2}\mathbb{E}[V_2(\theta_0)] + 3\mathrm{v}_3(\overline{M}_2) + c_1(\lambda_{\max} + a^{-1}) + 1,$$

where the process $\bar{\zeta}_t^{\lambda,n}$ is defined in Definition 1 and c_1 is given in (12). *Furthermore,*

$$\mathbb{E}[V_4(\bar{\zeta}_t^{\lambda,n})] \leq 2e^{-a\lambda t}\mathbb{E}[V_4(\theta_0)] + 3\mathrm{v}_5(\overline{M}_4) + 2c_3(\lambda_{\max} + a^{-1}) + 2,$$

where c_3 is given in (13).

Proof. For any $p \geq 1$, application of Ito's lemma and taking expectation yields

$$\mathbb{E}[V_p(\bar{\zeta}_t^{\lambda,n})] = \mathbb{E}[V_p(\bar{\theta}_{nT}^{\lambda})] + \int_{nT}^{t} \mathbb{E}\left[\lambda \frac{\Delta V_p(\bar{\zeta}_s^{\lambda,n})}{\beta} - \lambda \langle h(\bar{\zeta}_s^{\lambda,n}), \nabla V_p(\bar{\zeta}_s^{\lambda,n}) \rangle\right] ds.$$

Differentiating both sides and using Lemma 2, we arrive at

$$\frac{d}{dt}\mathbb{E}[V_p(\bar{\zeta}_t^{\lambda,n})] = \mathbb{E}\left[\lambda \frac{\Delta V_p(\bar{\zeta}_t^{\lambda,n})}{\beta} - \lambda \langle h(\bar{\zeta}_t^{\lambda,n}), \nabla V_p(\bar{\zeta}_t^{\lambda,n}) \rangle\right] \leq -\lambda \bar{c}(p)\mathbb{E}[V_p(\bar{\zeta}_t^{\lambda,n})] + \lambda \tilde{c}(p),$$

which yields

$$\begin{aligned}
\mathbb{E}[V_p(\bar{\zeta}_t^{\lambda,n})] &\leq e^{-\lambda(t-nT)\bar{c}(p)}\mathbb{E}[V_p(\bar{\theta}_{nT}^{\lambda})] + \frac{\tilde{c}(p)}{\bar{c}(p)}\left(1 - e^{-\lambda \bar{c}(p)(t-nT)}\right) \\
&\leq e^{-\lambda(t-nT)\bar{c}(p)}\mathbb{E}[V_p(\bar{\theta}_{nT}^{\lambda})] + \frac{\tilde{c}(p)}{\bar{c}(p)}.
\end{aligned}$$

Now for $p = 2$, by using Corollary 5,one obtains

$$\mathbb{E}[V_2(\bar{\zeta}_t^{\lambda,n})] \leq e^{-\lambda(t-nT)\bar{c}(2)}\mathbb{E}[V_2(\bar{\theta}_{nT}^{\lambda})] + \frac{\tilde{c}(2)}{\bar{c}(2)}$$

$$
\begin{aligned}
&\le (1-a\lambda)^{nT} e^{-\lambda(t-nT)\bar{c}(2)} \mathbb{E}[V_2(\theta_0)] + \frac{\tilde{c}(2)}{\bar{c}(2)} + c_1(\lambda_{\max} + a^{-1}) + 1 \\
&\le e^{-a\lambda t/2} \mathbb{E}[V_2(\theta_0)] + 3\mathrm{v}_3(\overline{M}_2) + c_1(\lambda_{\max} + a^{-1}) + 1,
\end{aligned}
$$

where the last inequality holds due to $1 - z \le e^{-z}$ for $z \ge 0$ and $\bar{c}(2) = a/2$. Similarly, for $p = 4$, one obtains

$$
\begin{aligned}
\mathbb{E}[V_4(\bar{\zeta}_t^{\lambda,n})] &\le e^{-\lambda(t-nT)\bar{c}(4)} \mathbb{E}[V_4(\bar{\theta}_{nT}^{\lambda})] + \frac{\tilde{c}(4)}{\bar{c}(4)} \\
&\le 2(1-a\lambda)^{nT} e^{-\lambda(t-nT)\bar{c}(4)} \mathbb{E}[V_4(\theta_0)] + \frac{\tilde{c}(4)}{\bar{c}(4)} + 2c_3(\lambda_{\max} + a^{-1}) + 2 \\
&\le 2e^{-a\lambda t} \mathbb{E}[V_4(\theta_0)] + 3\mathrm{v}_5(\overline{M}_4) + 2c_3(\lambda_{\max} + a^{-1}) + 2,
\end{aligned}
$$

where the last inequality holds due to $1 - z \le e^{-z}$ for $z \ge 0$ and $\bar{c}(4) = a$. □

3.3 Proof of the main theorems

We introduce a functional which is crucial to obtain the convergence rate in W_1. For any $p \ge 1$, $\mu, \nu \in \mathcal{P}_{V_p}$,

$$
w_{1,p}(\mu, \nu) := \inf_{\zeta \in \mathcal{C}(\mu,\nu)} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} [1 \wedge |\theta - \theta'|](1 + V_p(\theta) + V_p(\theta'))\zeta(d\theta d\theta'), \tag{20}
$$

and it satisfies trivially

$$
W_1(\mu, \nu) \le w_{1,p}(\mu, \nu). \tag{21}
$$

The case $p = 2$, i.e. $w_{1,2}$, is used throughout the section. The result below states a contraction property of $w_{1,2}$.

Proposition 1. *Let Z'_t, $t \in \mathbb{R}_+$ be the solution of* (3) *with initial condition $Z'_0 = \theta_0$ which is independent of $\mathcal{F}_\infty$ and satisfies $|\theta_0|_2$ is finite. Then,*

$$
w_{1,2}(\mathcal{L}(Z_t), \mathcal{L}(Z'_t)) \le \hat{c} e^{-\dot{c}t} w_{1,2}(\mathcal{L}(\theta_0), \mathcal{L}(\theta'_0)),
$$

where the constants $\dot{c}$ and $\hat{c}$ are given in Lemma 7.

Proof. See Proposition 3.14 of [8]. □

By using the contraction property provided in Proposition 1, one can construct the non-asymptotic bound between $\mathcal{L}(\bar{\theta}_t^{\lambda})$ and $\mathcal{L}(Z_t^{\lambda})$, $t \in [nT, (n+1)T]$, in W_1 distance by decomposing the error using the auxiliary process $\bar{\zeta}_t^{\lambda,n}$:

$$
W_1(\mathcal{L}(\bar{\theta}_t^{\lambda}), \mathcal{L}(Z_t^{\lambda})) \le W_1(\mathcal{L}(\bar{\theta}_t^{\lambda}), \mathcal{L}(\bar{\zeta}_t^{\lambda,n})) + W_1(\mathcal{L}(\bar{\zeta}_t^{\lambda,n}), \mathcal{L}(Z_t^{\lambda})). \tag{22}
$$

One notices that when $1 < \lambda \le \lambda_{\max}$, the result holds trivially. Thus, we consider the case $0 < \lambda \le 1$, which implies $1/2 < \lambda T \le 1$.

An upper bound for the first term in (22) is obtained in the Lemma below.

Lemma 4. *Let Assumption 1, 2, 3, and 4 hold. For any $0 < \lambda < \lambda_{\max}$ given in* (7)*, $t \in [nT, (n+1)T]$,*

$$
W_1(\mathcal{L}(\bar{\theta}_t^{\lambda}), \mathcal{L}(\bar{\zeta}_t^{\lambda,n})) \le \sqrt{\lambda}(e^{-an/2} \bar{C}_{2,1} \mathbb{E}[V_2(\theta_0)] + \bar{C}_{2,2})^{1/2},
$$

where $\bar{C}_{2,1}$ and $\bar{C}_{2,2}$ are given in (25).

Proof. To handle the first term in (22), we start by establishing an upper bound in Wasserstein-2 distance and the statement follows by noticing $W_1 \le W_2$. By employing synchronous coupling, using (11) and the definition of $\bar{\zeta}_t^{\lambda,n}$ in Definition 1, one obtains

$$
\left|\bar{\zeta}_t^{\lambda,n} - \bar{\theta}_t^{\lambda}\right| \le \lambda \left| \int_{nT}^{t} \left[H(\bar{\theta}_{\lfloor s \rfloor}^{\lambda}, X_{\lceil s \rceil}) - h(\bar{\zeta}_s^{\lambda,n}) \right] ds \right|.
$$

Then, the triangle inequality leads

$$\left|\bar{\zeta}_t^{\lambda,n} - \bar{\theta}_t^{\lambda}\right| \le \lambda \left| \int_{nT}^{t} \left[H(\bar{\theta}_{\lfloor s \rfloor}^{\lambda}, X_{\lceil s \rceil}) - h(\bar{\theta}_{\lfloor s \rfloor}^{\lambda}) \right] ds \right| + \lambda \left| \int_{nT}^{t} \left[h(\bar{\theta}_{\lfloor s \rfloor}^{\lambda}) - h(\bar{\zeta}_s^{\lambda,n}) \right] ds \right|.$$

Taking squares on both sides and the application of Remark 2 yield

$$\left|\bar{\zeta}_t^{\lambda,n} - \bar{\theta}_t^{\lambda}\right|^2 \le 2\lambda^2 \left| \int_{nT}^{t} \left[H(\bar{\theta}_{\lfloor s \rfloor}^{\lambda}, X_{\lceil s \rceil}) - h(\bar{\theta}_{\lfloor s \rfloor}^{\lambda}) \right] ds \right|^2 + 2\lambda L^2 \int_{nT}^{t} \left| \bar{\theta}_{\lfloor s \rfloor}^{\lambda} - \bar{\zeta}_s^{\lambda,n} \right|^2 ds.$$

By taking expectations on both sides and by using $(a+b)^2 \le 2a^2 + 2b^2$, for $a, b > 0$, one obtains

$$\begin{aligned}
\mathbb{E}\left[\left|\bar{\zeta}_t^{\lambda,n} - \bar{\theta}_t^{\lambda}\right|^2\right] &\le 2\lambda^2 \mathbb{E}\left[\left| \int_{nT}^{t} \left[H(\bar{\theta}_{\lfloor s \rfloor}^{\lambda}, X_{\lceil s \rceil}) - h(\bar{\theta}_{\lfloor s \rfloor}^{\lambda}) \right] ds \right|^2 \right] \\
&\quad + 4\lambda L^2 \int_{nT}^{t} \mathbb{E}\left[\left| \bar{\theta}_{\lfloor s \rfloor}^{\lambda} - \bar{\theta}_s^{\lambda} \right|^2 \right] ds + 4\lambda L^2 \int_{nT}^{t} \mathbb{E}\left[\left| \bar{\theta}_s^{\lambda} - \bar{\zeta}_s^{\lambda,n} \right|^2 \right] ds,
\end{aligned}$$

which implies due to $\lambda T \le 1$ and Lemma 13

$$\begin{aligned}
\mathbb{E}\left[\left|\bar{\zeta}_t^{\lambda,n} - \bar{\theta}_t^{\lambda}\right|^2\right] &\le 4\lambda L^2 (e^{-a\lambda nT} \bar{\sigma}_Y \mathbb{E}[V_2(\theta_0)] + \tilde{\sigma}_Y) + 4\lambda L^2 \int_{nT}^{t} \mathbb{E}\left[\left| \bar{\theta}_s^{\lambda} - \bar{\zeta}_s^{\lambda,n} \right|^2 \right] ds \\
&\quad + 2\lambda^2 \mathbb{E}\left[\left| \int_{nT}^{t} \left[H(\bar{\theta}_{\lfloor s \rfloor}^{\lambda}, X_{\lceil s \rceil}) - h(\bar{\theta}_{\lfloor s \rfloor}^{\lambda}) \right] ds \right|^2 \right],
\end{aligned} \tag{23}$$

where $\bar{\sigma}_Y$ and $\tilde{\sigma}_Y$ are provided in (52). Next, we bound the last term in (23) by partitioning the integral. Assume that $nT + K \le t \le nT + K + 1$ where $K + 1 \le T$. Thus we can write

$$\left| \int_{nT}^{t} \left[H(\bar{\theta}_{\lfloor s \rfloor}^{\lambda}, X_{\lceil s \rceil}) - h(\bar{\theta}_{\lfloor s \rfloor}^{\lambda}) \right] ds \right| = \left| \sum_{k=1}^{K} I_k + R_K \right|$$

where

$$I_k = H(\bar{\theta}_{nT+k-1}^{\lambda}, X_{nT+k}) - h(\bar{\theta}_{nT+k-1}^{\lambda}), \quad R_K = (t - (nT+K))(H(\bar{\theta}_{nT+K}^{\lambda}, X_{nT+K+1}) - h(\bar{\theta}_{nT+K}^{\lambda})).$$

Taking squares of both sides

$$\left| \sum_{k=1}^{K} I_k + R_K \right|^2 = \sum_{k=1}^{K} |I_k|^2 + 2 \sum_{k=2}^{K} \sum_{j=1}^{k-1} \langle I_k, I_j \rangle + 2 \sum_{k=1}^{K} \langle I_k, R_K \rangle + |R_K|^2,$$

Finally, we take expectations of both sides. Define the filtration $\mathcal{H}_t = \mathcal{F}_\infty^\lambda \vee \mathcal{G}_{\lfloor t \rfloor}$. We first note that for any $k = 2, \dots, K$, $j = 1, \dots, k-1$,

$$\begin{aligned}
&\mathbb{E}\langle I_k, I_j \rangle \\
&= \mathbb{E}\left[\mathbb{E}[\langle I_k, I_j \rangle | \mathcal{H}_{nT+k-1}] \right], \\
&= \mathbb{E}\left[\mathbb{E}\left[\left\langle H(\bar{\theta}_{nT+k-1}^{\lambda}, X_{nT+k}) - h(\bar{\theta}_{nT+k-1}^{\lambda}), H(\bar{\theta}_{nT+j-1}^{\lambda}, X_{nT+j}) - h(\bar{\theta}_{nT+j-1}^{\lambda}) \right\rangle \middle| \mathcal{H}_{nT+k-1} \right] \right], \\
&= \mathbb{E}\left[\left\langle \mathbb{E}\left[H(\bar{\theta}_{nT+k-1}^{\lambda}, X_{nT+k}) - h(\bar{\theta}_{nT+k-1}^{\lambda}) \middle| \mathcal{H}_{nT+k-1} \right], H(\bar{\theta}_{nT+j-1}^{\lambda}, X_{nT+j}) - h(\bar{\theta}_{nT+j-1}^{\lambda}) \right\rangle \right], \\
&= 0.
\end{aligned}$$

By the same argument $\mathbb{E}\langle I_k, R_K \rangle = 0$ for all $1 \le k \le K$. Therefore, the last term of (23) is bounded as

$$2\lambda^2 \mathbb{E}\left[\left| \int_{nT}^{t} \left[H(\bar{\theta}_{\lfloor s \rfloor}^{\lambda}, X_{\lceil s \rceil}) - h(\bar{\theta}_{\lfloor s \rfloor}^{\lambda}) \right] ds \right|^2 \right] = 2\lambda^2 \sum_{k=1}^{K} \mathbb{E}\left[|I_k|^2 \right] + 2\lambda^2 \mathbb{E}\left[|R_K|^2 \right]$$

$$\leq 2\lambda(e^{-a\lambda nT}\bar{\sigma}_Z\mathbb{E}[V_2(\theta_0)] + \tilde{\sigma}_Z),$$

where the last inequality holds due to Lemma 12 and $\bar{\sigma}_Z$ and $\tilde{\sigma}_Z$ are provided in (51). Therefore, the bound (23) becomes

$$\mathbb{E}\left[\left|\bar{\zeta}_t^{\lambda,n} - \bar{\theta}_t^\lambda\right|^2\right] \leq 4\lambda L^2 \int_{nT}^t \mathbb{E}\left[\left|\bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}\right|^2\right] ds$$
$$+ 4\lambda e^{-a\lambda nT}(L^2\bar{\sigma}_Y + \bar{\sigma}_Z)\mathbb{E}[V_2(\theta_0)] + 4\lambda(L^2\tilde{\sigma}_Y + \tilde{\sigma}_Z),$$

Using Grönwall's inequality yields

$$\mathbb{E}\left[\left|\bar{\zeta}_t^{\lambda,n} - \bar{\theta}_t^\lambda\right|^2\right] \leq 4\lambda e^{4L^2}\left[e^{-a\lambda nT}(L^2\bar{\sigma}_Y + \bar{\sigma}_Z)\mathbb{E}[V_2(\theta_0)] + (L^2\tilde{\sigma}_Y + \tilde{\sigma}_Z)\right],$$

which implies by $\lambda T \geq 1/2$,

$$W_2^2(\mathcal{L}(\bar{\theta}_t^\lambda), \mathcal{L}(\bar{\zeta}_t^{\lambda,n})) \leq \mathbb{E}\left[\left|\bar{\zeta}_t^{\lambda,n} - \bar{\theta}_t^\lambda\right|^2\right] \leq \lambda(e^{-an/2}\bar{C}_{2,1}\mathbb{E}[V_2(\theta_0)] + \bar{C}_{2,2}), \tag{24}$$

where

$$\bar{C}_{2,1} = 4e^{4L^2}(L^2\bar{\sigma}_Y + \bar{\sigma}_Z), \quad \bar{C}_{2,2} = 4e^{4L^2}(L^2\tilde{\sigma}_Y + \tilde{\sigma}_Z) \tag{25}$$

with $\bar{\sigma}_Y$, $\tilde{\sigma}_Y$ provided in (52) and $\bar{\sigma}_Z$, $\tilde{\sigma}_Z$ given in (51). □

Then, the following Lemma provides the bound for the second term in (22).

Lemma 5. *Let Assumption 1, 2, 3 and 4 hold. For any* $0 < \lambda < \lambda_{\max}$ *given in* (7), $t \in [nT, (n+1)T]$,

$$W_1(\mathcal{L}(\bar{\zeta}_t^{\lambda,n}), \mathcal{L}(Z_t^\lambda)) \leq \sqrt{\lambda}(e^{-\min\{\dot{c},a/2\}n/2}\bar{C}_{2,3}\mathbb{E}[V_4(\theta_0)] + \bar{C}_{2,4}),$$

where $\bar{C}_{2,3}$, $\bar{C}_{2,4}$ *is given in* (26).

Proof. To upper bound the second term $W_1(\mathcal{L}(\bar{\zeta}_t^{\lambda,n}), \mathcal{L}(Z_t^\lambda))$ in (22), we adapt the proof from Lemma 3.28 in [8]. By Proposition 1, Corollary 5, Lemma 3 and 4, one obtains

$$\begin{aligned}
&W_1(\mathcal{L}(\bar{\zeta}_t^{\lambda,n}), \mathcal{L}(Z_t^\lambda)) \\
&\leq \sum_{k=1}^n W_1(\mathcal{L}(\bar{\zeta}_t^{\lambda,k}), \mathcal{L}(\bar{\zeta}_t^{\lambda,k-1})), \\
&\leq \sum_{k=1}^n w_{1,2}(\mathcal{L}(\zeta_t^{kT,\bar{\theta}_{kT}^\lambda,\lambda}), \mathcal{L}(\zeta_t^{kT,\bar{\zeta}_{kT}^{\lambda,k-1},\lambda})) \\
&\leq \hat{c}\sum_{k=1}^n \exp(-\dot{c}(n-k))w_{1,2}(\mathcal{L}(\bar{\theta}_{kT}^\lambda), \mathcal{L}(\bar{\zeta}_{kT}^{\lambda,k-1})) \\
&\leq \hat{c}\sum_{k=1}^n \exp(-\dot{c}(n-k))W_2(\mathcal{L}(\bar{\theta}_{kT}^\lambda), \mathcal{L}(\bar{\zeta}_{kT}^{\lambda,k-1}))\left[1 + \left\{\mathbb{E}[V_4(\bar{\theta}_{kT}^\lambda)]\right\}^{1/2} + \left\{\mathbb{E}[V_4(\bar{\zeta}_{kT}^{\lambda,k-1})]\right\}^{1/2}\right] \\
&\leq (\sqrt{\lambda})^{-1}\hat{c}\sum_{k=1}^n \exp(-\dot{c}(n-k))W_2^2(\mathcal{L}(\bar{\theta}_{kT}^\lambda), \mathcal{L}(\bar{\zeta}_{kT}^{\lambda,k-1})) \\
&\quad + 3\sqrt{\lambda}\hat{c}\sum_{k=1}^n \exp(-\dot{c}(n-k))\left[1 + \mathbb{E}[V_4(\bar{\theta}_{kT}^\lambda)] + \mathbb{E}[V_4(\bar{\zeta}_{kT}^{\lambda,k-1})]\right] \\
&\leq \sqrt{\lambda}e^{-\min\{\dot{c},a/2\}n}n\hat{c}(e^{\min\{\dot{c},a/2\}}\bar{C}_{2,1}\mathbb{E}[V_2(\theta_0)] + 12\mathbb{E}[V_4(\theta_0)]) \\
&\quad + \sqrt{\lambda}\frac{\hat{c}}{1-\exp(-\dot{c})}(\bar{C}_{2,2} + 12c_3(\lambda_{\max} + a^{-1}) + 9\mathrm{v}_5(\overline{M}_4) + 15) \\
&\leq \sqrt{\lambda}(e^{-\min\{\dot{c},a/2\}n/2}\bar{C}_{2,3}\mathbb{E}[V_4(\theta_0)] + \bar{C}_{2,4})
\end{aligned}$$

where the last inequality holds due to $e^{-\alpha n}(n+1) \leq 1+\alpha^{-1}$, for $\alpha > 0$, and we take $\alpha = \min\{\dot{c}, a/2\}/2$, moreover,

$$\begin{aligned}
\bar{C}_{2,3} &= \hat{c}\left(1 + \frac{2}{\min\{\dot{c}, a/2\}}\right)(e^{\min\{\dot{c},a/2\}}\bar{C}_{2,1} + 12) \\
\bar{C}_{2,4} &= \frac{\hat{c}}{1-\exp(-\dot{c})}(\bar{C}_{2,2} + 12c_3(\lambda_{\max} + a^{-1}) + 9\mathrm{v}_5(\overline{M}_4) + 15)
\end{aligned} \tag{26}$$

with $\bar{C}_{2,1}$, $\bar{C}_{2,2}$ given in 25, $\hat{c}$, $\dot{c}$ given in Lemma 7, c_3 is given in (13) and $\overline{M}_4$ given in (19). □

By using similar arguments as in Lemma 5, an analogous result can be obtained in W_2 distance, which is given in the following corollary.

Corollary 6. *Let Assumption 1, 2, 3 and 4 hold. For any* $0 < \lambda < \lambda_{\max}$ *given in* (7), $t \in [nT, (n+1)T]$,

$$W_2(\mathcal{L}(\bar{\zeta}_t^{\lambda,n}), \mathcal{L}(Z_t^{\lambda})) \leq \lambda^{1/4}(e^{-\min\{\dot{c},a/2\}n/4}\bar{C}_{2,3}^{*}\mathbb{E}^{1/2}[V_4(\theta_0)] + \bar{C}_{2,4}^{*}),$$

where $\bar{C}_{2,3}^{*}$, $\bar{C}_{2,4}^{*}$ *is given in* (27).

Proof. One notices that $W_2 \leq \sqrt{2w_{1,2}}$, then one writes

$$\begin{aligned}
&W_2(\mathcal{L}(\bar{\zeta}_t^{\lambda,n}), \mathcal{L}(Z_t^{\lambda})) \\
&\leq \sum_{k=1}^{n} W_2(\mathcal{L}(\bar{\zeta}_t^{\lambda,k}), \mathcal{L}(\bar{\zeta}_t^{\lambda,k-1})) \\
&\leq \sum_{k=1}^{n} \sqrt{2}w_{1,2}^{1/2}(\mathcal{L}(\zeta_t^{kT,\bar{\theta}_{kT}^{\lambda},\lambda}), \mathcal{L}(\zeta_t^{kT,\bar{\zeta}_{kT}^{\lambda,k-1},\lambda})) \\
&\leq \sqrt{2}\hat{c}\sum_{k=1}^{n} \exp(-\dot{c}(n-k)/2)W_2^{1/2}(\mathcal{L}(\bar{\theta}_{kT}^{\lambda}), \mathcal{L}(\bar{\zeta}_{kT}^{\lambda,k-1}))\left[1 + \left\{\mathbb{E}[V_4(\bar{\theta}_{kT}^{\lambda})]\right\}^{1/2} + \left\{\mathbb{E}[V_4(\bar{\zeta}_{kT}^{\lambda,k-1})]\right\}^{1/2}\right]^{1/2} \\
&\leq \lambda^{-1/4}\sqrt{2}\hat{c}\sum_{k=1}^{n} \exp(-\dot{c}(n-k)/2)W_2(\mathcal{L}(\bar{\theta}_{kT}^{\lambda}), \mathcal{L}(\bar{\zeta}_{kT}^{\lambda,k-1})) \\
&\quad + \lambda^{1/4}\sqrt{2}\hat{c}\sum_{k=1}^{n} \exp(-\dot{c}(n-k)/2)\left[1 + \left\{\mathbb{E}[V_4(\bar{\theta}_{kT}^{\lambda})]\right\}^{1/2} + \left\{\mathbb{E}[V_4(\bar{\zeta}_{kT}^{\lambda,k-1})]\right\}^{1/2}\right] \\
&\leq \sqrt{2}\hat{c}\lambda^{1/4}e^{-\min\{\dot{c},a/2\}n/2}n(e^{\min\{\dot{c},a/2\}/2}\bar{C}_{2,1}^{1/2}\mathbb{E}^{1/2}[V_2(\theta_0)] + 2\sqrt{2}\mathbb{E}^{1/2}[V_4(\theta_0)]) \\
&\quad + \sqrt{2}\hat{c}\lambda^{1/4}\frac{1}{1-\exp(-\dot{c}/2)}(\bar{C}_{2,2}^{1/2} + 2\sqrt{2c_3}(\lambda_{\max} + a^{-1})^{1/2} + \sqrt{3}\mathrm{v}_5^{1/2}(\overline{M}_4) + \sqrt{15}) \\
&\leq \lambda^{1/4}(e^{-\min\{\dot{c},a/2\}n/4}\bar{C}_{2,3}^{*}\mathbb{E}^{1/2}[V_4(\theta_0)] + \bar{C}_{2,4}^{*}),
\end{aligned}$$

where

$$\begin{aligned}
\bar{C}_{2,3}^{*} &= \sqrt{2}\hat{c}\left(1 + \frac{4}{\min\{\dot{c}, a/2\}}\right)(e^{\min\{\dot{c},a/2\}/2}\bar{C}_{2,1}^{1/2} + 2\sqrt{2}) \\
\bar{C}_{2,4}^{*} &= \frac{\sqrt{2}\hat{c}}{1-\exp(-\dot{c}/2)}(\bar{C}_{2,2}^{1/2} + 2\sqrt{2c_3}(\lambda_{\max} + a^{-1})^{1/2} + \sqrt{3}\mathrm{v}_5^{1/2}(\overline{M}_4) + \sqrt{15}),
\end{aligned} \tag{27}$$

with $\bar{C}_{2,1}$, $\bar{C}_{2,2}$ given in 25, $\hat{c}$, $\dot{c}$ given in Lemma 7, c_3 is given in (13) and $\overline{M}_4$ given in Lemma 2. This completes the proof. □

Finally, by using the inequality (22) and the results from previous lemmas, one can obtain the non-asymptotic bound between $\bar{\theta}_t^{\lambda}$ and Z_t^{λ}, $t \in [nT, (n+1)T]$, in W_1 distance.

Lemma 6. *Let Assumption 1, 2, 3 and 4 hold. For any* $0 < \lambda < \lambda_{\max}$ *given in* (7), $t \in [nT, (n+1)T]$,

$$W_1(\mathcal{L}(\bar{\theta}_t^{\lambda}), \mathcal{L}(Z_t^{\lambda})) \leq \bar{C}_2\sqrt{\lambda}(e^{-\min\{\dot{c},a/2\}n/2}\mathbb{E}[V_4(\theta_0)] + 1),$$

where $\bar{C}_2$ *is given in* (28).

Proof. By using Lemma 4 and 5, one obtains

$$
\begin{aligned}
& W_1(\mathcal{L}(\bar{\theta}_t^\lambda), \mathcal{L}(Z_t^\lambda)) \\
& \le W_1(\mathcal{L}(\bar{\theta}_t^\lambda), \mathcal{L}(\bar{\zeta}_t^{\lambda,n})) + W_1(\mathcal{L}(\bar{\zeta}_t^{\lambda,n}), \mathcal{L}(Z_t^\lambda)) \\
& \le \sqrt{\lambda}(e^{-an/2}\bar{C}_{2,1}^{1/2}\mathbb{E}^{1/2}[V_2(\theta_0)] + \bar{C}_{2,2}^{1/2}) + \sqrt{\lambda}(e^{-\min\{\dot{c},a/2\}n/2}\bar{C}_{2,3}\mathbb{E}[V_4(\theta_0)] + \bar{C}_{2,4}) \\
& \le \bar{C}_2\sqrt{\lambda}(e^{-\min\{\dot{c},a/2\}n/2}\mathbb{E}[V_4(\theta_0)] + 1),
\end{aligned}
$$

where

$$
\bar{C}_2 = \bar{C}_{2,1}^{1/2} + \bar{C}_{2,2}^{1/2} + \bar{C}_{2,3} + \bar{C}_{2,4}. \tag{28}
$$

□

Before proceeding to the proofs of the main results, we provide explicitly the constants $\dot{c}$ and $\hat{c}$ in Proposition 1.

Lemma 7. *The contraction constant in Proposition 1 is given by*

$$
\dot{c} = \min\{\bar{\phi}, \bar{c}(p), 4\tilde{c}(p)\dot{\epsilon}\bar{c}(p)\}/2,
$$

where the explicit expressions for $\bar{c}(p)$ and $\tilde{c}(p)$ can be found in Lemma 2 and $\bar{\phi}$ is given by

$$
\bar{\phi} = \left(\sqrt{4\pi/L}\bar{b}\exp\left(\left(\bar{b}\sqrt{L}/2 + 2/\sqrt{L}\right)^2\right)\right)^{-1}.
$$

Furthermore, any $\dot{\epsilon}$ can be chosen which satisfies the following inequality

$$
\dot{\epsilon} \le 1 \wedge \left(8\tilde{c}(p)\sqrt{\pi/L}\int_0^{\tilde{b}} \exp\left(\left(s\sqrt{L}/2 + 2/\sqrt{L}\right)^2\right) ds\right)^{-1},
$$

where $\tilde{b} = \sqrt{2\tilde{c}(p)/\bar{c}(p) - 1}$ and $\bar{b} = \sqrt{4\tilde{c}(p)(1+\bar{c}(p))/\bar{c}(p) - 1}$. The constant $\hat{c}$ is given as the ratio C_{11}/C_{10}, where C_{11}, C_{10} are given explicitly in [8, Lemma 3.26].

Proof. See [8, Lemma 3.26]. □

Proof of Theorem 1 One notes that, by Lemma 6 and Proposition 1, for $t \in [nT, (n+1)T]$

$$
\begin{aligned}
W_1(\mathcal{L}(\bar{\theta}_t^\lambda), \pi_\beta) & \le W_1(\mathcal{L}(\bar{\theta}_t^\lambda), \mathcal{L}(Z_t^\lambda)) + W_1(\mathcal{L}(Z_t^\lambda), \pi_\beta) \\
& \le \bar{C}_2\sqrt{\lambda}(e^{-\min\{\dot{c},a/2\}n/2}\mathbb{E}[V_4(\theta_0)] + 1) + \hat{c}e^{-\dot{c}\lambda t}w_{1,2}(\theta_0, \pi_\beta) \\
& \le \bar{C}_2\sqrt{\lambda}(e^{-\min\{\dot{c},a/2\}n/2}\mathbb{E}[V_4(\theta_0)] + 1) + \hat{c}e^{-\dot{c}\lambda t}\left[1 + \mathbb{E}[V_2(\theta_0)] + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta)\right] \\
& \le 2e^{-\min\{\dot{c},a/2\}n/2}(\lambda_{\max}^{1/2}\bar{C}_2 + \hat{c})(1 + \mathbb{E}[|\theta_0|^4]) \\
& \quad + \hat{c}e^{-\min\{\dot{c},a/2\}n/2}\left[1 + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta)\right] + \sqrt{\lambda}\bar{C}_2,
\end{aligned}
$$

which implies, for any $n \in \mathbb{N}$

$$
W_1(\mathcal{L}(\theta_n^\lambda), \pi_\beta) \le C_1 e^{-C_0\lambda n}(1 + \mathbb{E}[|\theta_0|^4]) + C_2\sqrt{\lambda},
$$

where

$$
C_0 = \min\{\dot{c}, a/2\}/2, \quad C_1 = 2\left[(\lambda_{\max}^{1/2}\bar{C}_2 + \hat{c}) + \hat{c}\left(1 + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta)\right)\right], \quad C_2 = \bar{C}_2, \tag{29}
$$

with $\bar{C}_2$ given in 28.

Proof of Corollary 1 By using (24) in Lemma 4, Corollary 6 and Proposition 1, one obtains

$$
W_2(\mathcal{L}(\bar{\theta}_t^\lambda), \pi_\beta) \le W_2(\mathcal{L}(\bar{\theta}_t^\lambda), \mathcal{L}(Z_t^\lambda)) + W_2(\mathcal{L}(Z_t^\lambda), \pi_\beta)
$$

$$
\begin{aligned}
&\leq W_2(\mathcal{L}(\bar{\theta}_t^\lambda), \mathcal{L}(\bar{\zeta}_t^{\lambda,n})) + W_2(\mathcal{L}(\bar{\zeta}_t^{\lambda,n}), \mathcal{L}(Z_t^\lambda)) + W_2(\mathcal{L}(Z_t^\lambda), \pi_\beta) \\
&\leq \sqrt{\lambda}(e^{-an/2}\bar{C}_{2,1}\mathbb{E}[V_2(\theta_0)] + \bar{C}_{2,2})^{1/2} \\
&\quad + \lambda^{1/4}(e^{-\min\{\dot{c},a/2\}n/4}\bar{C}_{2,3}^*\mathbb{E}^{1/2}[V_4(\theta_0)] + \bar{C}_{2,4}^*), + \sqrt{2w_{1,2}(\mathcal{L}(Z_t^\lambda), \pi_\beta)} \\
&\leq \lambda^{1/4}\tilde{C}_2(e^{-\min\{\dot{c},a/2\}n/4}\mathbb{E}[V_4(\theta_0)] + 1) + \hat{c}^{1/2}e^{-\dot{c}\lambda t/2}\sqrt{2w_{1,2}(\theta_0, \pi_\beta)},
\end{aligned}
$$

where $\tilde{C}_2 = \lambda_{\max}^{1/4}\bar{C}_{2,1}^{1/2} + \lambda_{\max}^{1/4}\bar{C}_{2,2}^{1/2} + \bar{C}_{2,3}^* + \bar{C}_{2,4}^*$ and it can be further calculated as

$$
\begin{aligned}
W_2(\mathcal{L}(\bar{\theta}_t^\lambda), \pi_\beta) &\leq \lambda^{1/4}\tilde{C}_2(e^{-\min\{\dot{c},a/2\}n/4}\mathbb{E}[V_4(\theta_0)] + 1) \\
&\quad + \sqrt{2}\hat{c}^{1/2}e^{-\dot{c}\lambda t/2}\left(1 + \mathbb{E}[V_2(\theta_0)] + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta)\right)^{1/2} \\
&\leq 2e^{-\min\{\dot{c},a/2\}n/4}(\lambda_{\max}^{1/4}\tilde{C}_2 + \sqrt{2}\hat{c}^{1/2})(1 + \mathbb{E}[|\theta_0|^4]) \\
&\quad + \sqrt{2}\hat{c}^{1/2}e^{-\min\{\dot{c},a/2\}n/4}\left[1 + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta)\right] + \lambda^{1/4}\tilde{C}_2,
\end{aligned}
$$

Finally, one obtains

$$
W_2(\mathcal{L}(\theta_n^\lambda), \pi_\beta) \leq C_4 e^{-C_3\lambda n}\mathbb{E}[|\theta_0|^4 + 1] + C_5\lambda^{1/4}
$$

where

$$
C_3 = \min\{\dot{c}, a/2\}/4, \quad C_4 = 2\left[(\lambda_{\max}^{1/4}\tilde{C}_2 + \sqrt{2}\hat{c}^{1/2}) + \hat{c}^{1/2}\left(1 + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta)\right)\right], \quad C_5 = \tilde{C}_2. \tag{30}
$$

Proof of Corollary 2 To obtain an upper bound for the expected excess risk $\mathbb{E}[U(\hat{\theta})] - \inf_{\theta\in\mathbb{R}^d} U(\theta)$, one considers the following splitting

$$
\mathbb{E}[U(\hat{\theta})] - \inf_{\theta\in\mathbb{R}^d} U(\theta) = \left(\mathbb{E}[U(\hat{\theta})] - \mathbb{E}[U(Z_\infty)]\right) + \left(\mathbb{E}[U(Z_\infty)] - \inf_{\theta\in\mathbb{R}^d} U(\theta)\right), \tag{31}
$$

where $\hat{\theta} = \theta_n^\lambda$ and $Z_\infty \sim \pi_\beta$ with $\pi_\beta(\theta) = \exp(-\beta U(\theta))$ for all $\theta \in \mathbb{R}^d$. By using [19, Lemma 3.5], Lemma 1, 14 and Corollary 1, the first term on the RHS of (31) can be bounded by

$$
\begin{aligned}
&\mathbb{E}[U(\hat{\theta})] - \mathbb{E}[U(Z_\infty)] \\
&\leq \left(L(\mathbb{E}\left[|\theta_0|^2\right] + (c_1 + \mathbb{E}[K_1^2(X_0)]/a)(\lambda_{\max} + a^{-1}))^{1/2} + |h(0)|\right) W_2(\mathcal{L}(\theta_n^\lambda), \pi_\beta) \\
&\leq \left(L(\mathbb{E}\left[|\theta_0|^2\right] + (c_1 + \mathbb{E}[K_1^2(X_0)]/a)(\lambda_{\max} + a^{-1}))^{1/2} + |h(0)|\right)\left(C_4 e^{-C_3\lambda n}\mathbb{E}[|\theta_0|^4 + 1] + C_5\lambda^{1/4}\right) \\
&\leq \hat{C}_1 e^{-\hat{C}_0\lambda n} + \hat{C}_2\lambda^{1/4},
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{C}_0 &= C_3, \\
\hat{C}_1 &= C_4\left(L(\mathbb{E}\left[|\theta_0|^2\right] + (c_1 + \mathbb{E}[K_1^2(X_0)]/a)(\lambda_{\max} + a^{-1}))^{1/2} + |h(0)|\right)\mathbb{E}[|\theta_0|^4 + 1], \\
\hat{C}_2 &= C_5\left(L(\mathbb{E}\left[|\theta_0|^2\right] + (c_1 + \mathbb{E}[K_1^2(X_0)]/a)(\lambda_{\max} + a^{-1}))^{1/2} + |h(0)|\right),
\end{aligned} \tag{32}
$$

with C_3, C_4, C_5 given in (30) and c_1 given in (12). Moreover, the second term on the RHS of (31) can be estimated by using [19, Proposition 3.4], which gives,

$$
\mathbb{E}[U(Z_\infty)] - \inf_{\theta\in\mathbb{R}^d} U(\theta) \leq \frac{\hat{C}_3}{\beta},
$$

where

$$
\hat{C}_3 = \frac{d}{2}\log\left(\frac{e\beta L}{ad}\left(\frac{2d}{\beta} + 2b + \frac{\mathbb{E}[K_1^2(X_0)]}{a}\right)\right). \tag{33}
$$

Finally, one obtains

$$
\mathbb{E}[U(\hat{\theta})] - \inf_{\theta\in\mathbb{R}^d} U(\theta) \leq \hat{C}_1 e^{-\hat{C}_0\lambda n} + \hat{C}_2\lambda^{1/4} + \hat{C}_3/\beta.
$$

4 Proof of the main results: convex case

The analysis of the convergence results in the convex case, i.e. Theorem 2, relies on the properties of the LMC algorithm, known also as the unadjusted Langevin algorithm (ULA). The LMC algorithm associated with SDE (3) is given explicitly by, for any $n \in \mathbb{N}$,

$$\dot{\theta}^{\lambda}_{n+1} := \dot{\theta}^{\lambda}_{n} - \lambda h(\dot{\theta}^{\lambda}_{n}) + \sqrt{2\beta^{-1}\lambda}\xi_{n+1}, \quad \dot{\theta}^{\lambda}_{0} := \theta_0. \tag{34}$$

For $0 < \lambda < \bar{\lambda}_{\max}$, the Markov kernel $\dot{R}_\lambda$ associated with (34) is given by, for all $A \in \mathcal{B}(\mathbb{R}^d)$ and $\theta \in \mathbb{R}^d$,

$$\dot{R}_\lambda(\theta, A) = \int_A (4\beta^{-1}\pi\lambda)^{-d/2} \exp\left(-\beta(4\lambda)^{-1} \left|y - \theta + \lambda h(\theta)\right|^2\right) dy.$$

In this section, the moment estimates of the SDE (3), the LMC algorithm (34) and the SGLD algorithm (2) are presented which contribute to the analysis of the convergence resuts.

4.1 Preliminary estimates

Under Assumptions 5 and 6, U has a unique minimizer $\theta^* \in \mathbb{R}^d$. Denote by $(P_t)_{t\geq 0}$ the semigroup associated with SDE (3). The statements below provide a moment bound and a convergence result for SDE (3).

Lemma 8 (Proposition 1 in [12]). *Let Assumptions 1, 2, 3, 5 and 6 hold.*

(i) For all $t > 0$ and $y \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} |y - \theta^*|^2 P_t(\theta, dy) \leq |\theta - \theta^*|^2 e^{-2\hat{a}t} + (d/(\hat{a}\beta))(1 - e^{-2\hat{a}t}).$$

(ii) The stationary distribution π_β satisfies

$$\int_{\mathbb{R}^d} |y - \theta^*|^2 \pi_\beta(dy) \leq d/(\hat{a}\beta).$$

The following lemma provides moment estimates for $(\dot{\theta}_n)_{n\in\mathbb{N}}$ and it states that $\dot{R}_\lambda$ admits an invariant measure π_λ which may differ from π_β.

Lemma 9. *Let Assumptions 1, 2, 5 and 6 hold. Then, for all $0 < \lambda < \bar{\lambda}_{\max}$ given in (10), one obtains:*

(i) For all $t > 0$ and $\theta \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} |y - \theta^*|^2 \dot{R}^n_\lambda(\theta, dy) \leq (1 - 2\hat{a}^*\lambda)^n |\theta - \theta^*|^2 + (d/(\hat{a}^*\beta))(1 - (1 - 2\hat{a}^*\lambda)^n).$$

(ii) The Markov kernel $\dot{R}_\lambda$ has a unique stationary distribution π_λ and it satisfies

$$\int_{\mathbb{R}^d} |\theta - \theta^*|^2 \pi_\lambda(d\theta) \leq d/(\hat{a}^*\beta).$$

(iii) For all $n \in \mathbb{R}^d$ and $\theta \in \mathbb{R}^d$,

$$W_2(\delta_\theta \dot{R}^n_\lambda, \pi_\lambda) \leq e^{-\hat{a}^*\lambda n}(|\theta - \theta^*|^2 + d/(\hat{a}^*\beta))^{1/2}.$$

The lemma below presents a second moment bound for θ^λ_n in the convex case.

Lemma 10. *Let Assumptions 1, 2, 5 hold. For any* $0 < \lambda < \bar{\lambda}_{\max}$ *given in* (10),

$$\mathbb{E}\left[\left|\theta_n^\lambda - \theta^*\right|^2\right] \le (1-\hat{a}\lambda)^n \mathbb{E}\left[|\theta_0 - \theta^*|^2\right] + \bar{c}_4 \hat{a}^{-1},$$

where

$$\bar{c}_4 = 32\mathbb{E}\left[K_1^2(X_0)\right]\hat{a}_1^{-1} + 9\bar{\lambda}_{\max}(L_1^2 \mathbb{E}\left[K_\rho(X_0)\right]|\theta^*|^2 + L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^2(X_0)\right]) + 2d\beta^{-1}. \tag{35}$$

This implies $\sup_n \mathbb{E}\left[\left|\theta_{n+1}^\lambda - \theta^*\right|^2\right] \le \mathbb{E}\left[|\theta_0 - \theta^*|^2\right] + \bar{c}_4\hat{a}^{-1} < \infty$. *Furthermore, if* $\rho = 0$ *in Assumption 1, the result holds for* $\lambda \in \min\{1/2(\hat{a}+L), 1/(6L_1)\}$ *with* $\hat{a} = \hat{a}_1 + \hat{a}_2$.

Proof. By using (2), one writes, for any $n \in \mathbb{N}$,

$$\begin{aligned}
|\theta_{n+1}^\lambda - \theta^*|^2 &= |\theta_n^\lambda - \theta^*|^2 + 2\left\langle \theta_n^\lambda - \theta^*, -\lambda H(\theta_n^\lambda, X_{n+1}) + \sqrt{2\beta^{-1}\lambda}\xi_{n+1}\right\rangle \\
&\quad + |-\lambda H(\theta_n^\lambda, X_{n+1}) + \sqrt{2\beta^{-1}\lambda}\xi_{n+1}|^2 \\
&= |\theta_n^\lambda - \theta^*|^2 - 2\lambda\left\langle \theta_n^\lambda - \theta^*, H(\theta_n^\lambda, X_{n+1}) - H(\theta^*, X_{n+1})\right\rangle \\
&\quad - 2\lambda\left\langle \theta_n^\lambda - \theta^*, H(\theta^*, X_{n+1})\right\rangle + 2\left\langle \theta_n^\lambda - \theta^*, \sqrt{2\beta^{-1}\lambda}\xi_{n+1}\right\rangle \\
&\quad + \lambda^2|H(\theta_n^\lambda, X_{n+1})|^2 - 2\lambda\left\langle H(\theta_n^\lambda, X_{n+1}), \sqrt{2\beta^{-1}\lambda}\xi_{n+1}\right\rangle + 2\beta^{-1}\lambda|\xi_{n+1}|^2.
\end{aligned}$$

Taking conditional expectation on both sides and by using Remark 1 and Assumption 1, 5 yield

$$\begin{aligned}
&\mathbb{E}\left[|\theta_{n+1}^\lambda - \theta^*|^2\middle|\theta_n^\lambda\right] \\
&= |\theta_n^\lambda - \theta^*|^2 - 2\lambda\mathbb{E}\left[\left\langle \theta_n^\lambda - \theta^*, F(\theta_n^\lambda, X_{n+1}) - F(\theta^*, X_{n+1})\right\rangle\middle|\theta_n^\lambda\right] \\
&\quad - 2\lambda\mathbb{E}\left[\left\langle \theta_n^\lambda - \theta^*, G(\theta_n^\lambda, X_{n+1}) - G(\theta^*, X_{n+1})\right\rangle\middle|\theta_n^\lambda\right] \\
&\quad - 2\lambda\left\langle \theta_n^\lambda - \theta^*, h(\theta^*)\right\rangle + \lambda^2\mathbb{E}\left[|H(\theta_n^\lambda, X_{n+1})|^2\middle|\theta_n^\lambda\right] + 2d\beta^{-1}\lambda \\
&\le |\theta_n^\lambda - \theta^*|^2 - 2\lambda\hat{a}_1|\theta_n^\lambda - \theta^*|^2 + 4\lambda\mathbb{E}\left[K_1(X_0)\right]|\theta_n^\lambda - \theta^*| \\
&\quad + \lambda^2\mathbb{E}\left[\left((1+|X_{n+1}|)^{\rho+1}(L_1|\theta_n^\lambda - \theta^*| + L_1|\theta^*| + L_2) + F_*(X_{n+1})\right)^2\middle|\theta_n^\lambda\right] + 2d\beta^{-1}\lambda \\
&\le (1-2\hat{a}_1\lambda)|\theta_n^\lambda - \theta^*|^2 + 4\lambda\mathbb{E}\left[K_1(X_0)\right]|\theta_n^\lambda - \theta^*| + 2\lambda^2L_1^2\mathbb{E}\left[K_\rho(X_0)\right]|\theta_n^\lambda - \theta^*|^2 \\
&\quad + 6\lambda^2L_1^2\mathbb{E}\left[K_\rho(X_0)\right]|\theta^*|^2 + 6\lambda^2L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + 6\lambda^2\mathbb{E}\left[F_*^2(X_0)\right] + 2d\beta^{-1}\lambda
\end{aligned} \tag{36}$$

which implies, for $0 < \lambda < \bar{\lambda}_{\max}$,

$$\begin{aligned}
\mathbb{E}\left[|\theta_{n+1}^\lambda - \theta^*|^2\middle|\theta_n^\lambda\right] &\le \left(1 - \frac{3}{2}\hat{a}_1\lambda\right)|\theta_n^\lambda - \theta^*|^2 + 4\lambda\mathbb{E}\left[K_1(X_0)\right]|\theta_n^\lambda - \theta^*| \\
&\quad + 6\lambda^2L_1^2\mathbb{E}\left[K_\rho(X_0)\right]|\theta^*|^2 + 6\lambda^2L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + 6\lambda^2\mathbb{E}\left[F_*^2(X_0)\right] + 2d\beta^{-1}\lambda.
\end{aligned}$$

Then, for $|\theta_n^\lambda - \theta^*| > 8\mathbb{E}\left[K_1(X_0)\right]\hat{a}_1^{-1}$, one notices that

$$-\frac{1}{2}\hat{a}_1\lambda|\theta_n^\lambda - \theta^*|^2 + 4\lambda\mathbb{E}\left[K_1(X_0)\right]|\theta_n^\lambda - \theta^*| < 0,$$

and this indicates

$$\begin{aligned}
\mathbb{E}\left[|\theta_{n+1}^\lambda - \theta^*|^2\middle|\theta_n^\lambda\right] &\le (1-\hat{a}_1\lambda)|\theta_n^\lambda - \theta^*|^2 + 6\lambda^2L_1^2\mathbb{E}\left[K_\rho(X_0)\right]|\theta^*|^2 \\
&\quad + 6\lambda^2L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + 6\lambda^2\mathbb{E}\left[F_*^2(X_0)\right] + 2d\beta^{-1}\lambda.
\end{aligned}$$

Similarly, for $|\theta_n^\lambda - \theta^*| \le 8\mathbb{E}\left[K_1(X_0)\right]\hat{a}_1^{-1}$, one obtains

$$\mathbb{E}\left[|\theta_{n+1}^\lambda - \theta^*|^2\middle|\theta_n^\lambda\right] \le \left(1 - \frac{3}{2}\hat{a}_1\lambda\right)|\theta_n^\lambda - \theta^*|^2 + 32\lambda\mathbb{E}\left[K_1^2(X_0)\right]\hat{a}_1^{-1}$$

$$+ 6\lambda^2 L_1^2 \mathbb{E}\left[K_\rho(X_0)\right] |\theta^*|^2 + 6\lambda^2 L_2^2 \mathbb{E}\left[K_\rho(X_0)\right] + 6\lambda^2 \mathbb{E}\left[F_*^2(X_0)\right] + 2d\beta^{-1}\lambda.$$

Combining the two cases yields

$$\mathbb{E}\left[|\theta_{n+1}^\lambda - \theta^*|^2 \middle| \theta_n^\lambda \right] \leq (1 - \hat{a}\lambda)|\theta_n^\lambda - \theta^*|^2 + \lambda c_4,$$

where $c_4 = 32\mathbb{E}\left[K_1^2(X_0)\right] \hat{a}_1^{-1} + 6\bar{\lambda}_{\max}(L_1^2 \mathbb{E}\left[K_\rho(X_0)\right] |\theta^*|^2 + L_2^2 \mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^2(X_0)\right]) + 2d\beta^{-1}$. The result follows by induction.

Moreover, one observes that when $\rho = 0$ in Assumption 1, F is co-coercive, i.e. for any $\theta, \theta' \in \mathbb{R}^d$ and for every $x \in \mathbb{R}^m$

$$\left\langle \theta - \theta', F(\theta, x) - F(\theta', x) \right\rangle \geq \frac{1}{L_1} |F(\theta, x) - F(\theta', x)|^2. \tag{37}$$

Then, by substituting (37) into (36), one obtains

$$\begin{aligned}
\mathbb{E}\left[|\theta_{n+1}^\lambda - \theta^*|^2 \middle| \theta_n^\lambda \right] &\leq |\theta_n^\lambda - \theta^*|^2 - \frac{3}{2}\lambda \hat{a}_1 |\theta_n^\lambda - \theta^*|^2 - \frac{\lambda}{2L_1} \mathbb{E}\left[|F(\theta_n^\lambda, X_{n+1}) - F(\theta^*, X_{n+1})|^2 \middle| \theta_n^\lambda \right] \\
&\quad + 4\lambda \mathbb{E}\left[K_1(X_0)\right] |\theta_n^\lambda - \theta^*| + \lambda^2 \mathbb{E}\left[|H(\theta_n^\lambda, X_{n+1})|^2 \middle| \theta_n^\lambda \right] + 2d\beta^{-1}\lambda \\
&\leq \left(1 - \frac{3}{2}\lambda \hat{a}_1\right) |\theta_n^\lambda - \theta^*|^2 + 4\lambda \mathbb{E}\left[K_1(X_0)\right] |\theta_n^\lambda - \theta^*| \\
&\quad + \left(3\lambda^2 - \frac{\lambda}{2L_1}\right) \mathbb{E}\left[|F(\theta_n^\lambda, X_{n+1}) - F(\theta^*, X_{n+1})|^2 \middle| \theta_n^\lambda \right] \\
&\quad + 3\lambda^2 \mathbb{E}\left[|F(\theta^*, X_{n+1})|^2 \middle| \theta_n^\lambda \right] + 3\lambda^2 \mathbb{E}\left[K_1^2(X_0)\right] + 2d\beta^{-1}\lambda,
\end{aligned}$$

which implies for $\lambda \in \min\{1/2\hat{a}_1, 1/(6L_1)\}$

$$\begin{aligned}
\mathbb{E}\left[|\theta_{n+1}^\lambda - \theta^*|^2 \middle| \theta_n^\lambda \right] &\leq \left(1 - \frac{3}{2}\lambda \hat{a}_1\right) |\theta_n^\lambda - \theta^*|^2 + 4\lambda \mathbb{E}\left[K_1(X_0)\right] |\theta_n^\lambda - \theta^*| \\
&\quad + 9\lambda^2 L_1^2 \mathbb{E}\left[K_\rho(X_0)\right] |\theta^*|^2 + 9\lambda^2 L_2^2 \mathbb{E}\left[K_\rho(X_0)\right] + 9\lambda^2 \mathbb{E}\left[F_*^2(X_0)\right] + 2d\beta^{-1}\lambda.
\end{aligned}$$

By using the same arguments as above, consider the case $|\theta_n^\lambda - \theta^*| > 8\mathbb{E}\left[K_1(X_0)\right] \hat{a}_1^{-1}$, one notices that

$$-\frac{1}{2}\hat{a}_1 \lambda |\theta_n^\lambda - \theta^*|^2 + 4\lambda \mathbb{E}\left[K_1(X_0)\right] |\theta_n^\lambda - \theta^*| < 0,$$

and this indicates

$$\begin{aligned}
\mathbb{E}\left[|\theta_{n+1}^\lambda - \theta^*|^2 \middle| \theta_n^\lambda \right] &\leq (1 - \hat{a}_1 \lambda) |\theta_n^\lambda - \theta^*|^2 + 9\lambda^2 L_1^2 \mathbb{E}\left[K_\rho(X_0)\right] |\theta^*|^2 \\
&\quad + 9\lambda^2 L_2^2 \mathbb{E}\left[K_\rho(X_0)\right] + 9\lambda^2 \mathbb{E}\left[F_*^2(X_0)\right] + 2d\beta^{-1}\lambda.
\end{aligned}$$

Similarly, for $|\theta_n^\lambda - \theta^*| \leq 8\mathbb{E}\left[K_1(X_0)\right] \hat{a}_1^{-1}$, one obtains

$$\begin{aligned}
\mathbb{E}\left[|\theta_{n+1}^\lambda - \theta^*|^2 \middle| \theta_n^\lambda \right] &\leq \left(1 - \frac{3}{2}\hat{a}_1 \lambda\right) |\theta_n^\lambda - \theta^*|^2 + 32\lambda \mathbb{E}\left[K_1^2(X_0)\right] \hat{a}_1^{-1} \\
&\quad + 9\lambda^2 L_1^2 \mathbb{E}\left[K_\rho(X_0)\right] |\theta^*|^2 + 9\lambda^2 L_2^2 \mathbb{E}\left[K_\rho(X_0)\right] + 9\lambda^2 \mathbb{E}\left[F_*^2(X_0)\right] + 2d\beta^{-1}\lambda.
\end{aligned}$$

Combining the two cases yields

$$\mathbb{E}\left[|\theta_{n+1}^\lambda - \theta^*|^2 \middle| \theta_n^\lambda \right] \leq (1 - \hat{a}\lambda)|\theta_n^\lambda - \theta^*|^2 + \lambda \bar{c}_4,$$

where $\bar{c}_4 = 32\mathbb{E}\left[K_1^2(X_0)\right] \hat{a}_1^{-1} + 9\bar{\lambda}_{\max}(L_1^2 \mathbb{E}\left[K_\rho(X_0)\right] |\theta^*|^2 + L_2^2 \mathbb{E}\left[K_\rho(X_0)\right] + \mathbb{E}\left[F_*^2(X_0)\right]) + 2d\beta^{-1}$. $\square$

4.2 Convergence results

We aim to establish the non-asymptotic bound in Wasserstein-2 distance between $\mathcal{L}(\theta_n^\lambda)$ and π_β. To achieve this, we consider the following decomposition:

$$W_2(\mathcal{L}(\theta_n^\lambda), \pi_\beta) \leq W_2(\mathcal{L}(\theta_n^\lambda), \mathcal{L}(\dot{\theta}_n^\lambda)) + W_2(\mathcal{L}(\dot{\theta}_n^\lambda), \pi_\lambda) + W_2(\pi_\lambda, \pi_\beta). \tag{38}$$

The lemma presented below provides the non-asymptotic estimates for the last two terms in (38).

Theorem 3. *[12, Corollary 7] Let Assumptions 1, 2, 3, 5 and 6 hold. Then, for any* $0 < \lambda < \bar{\lambda}_{\max}$ *given in* (10)*, the Markov chain* $(\dot{\theta}_n^\lambda)_{n\in\mathbb{N}}$ *admits an invariant measure* π_λ *such that, for all* $n \in \mathbb{N}$*,*

$$W_2(\mathcal{L}(\dot{\theta}_n^\lambda), \pi_\lambda) \leq \bar{C}_7 e^{-\hat{a}^*\lambda n},$$

where $\bar{C}_7 = (|\theta_0 - \theta|^2 + d/\hat{a}^*\beta)^{1/2}$ *is given in Lemma 9 (iii) with* $\hat{a}^* = \hat{a}L/(\hat{a}+L)$*. Furthermore,*

$$W_2(\pi_\beta, \pi_\lambda) \leq \bar{C}_{8,1}\sqrt{\lambda},$$

where

$$\bar{C}_{8,1} = \left(dL^2(\hat{a}^*\beta)^{-1}(2\lambda + (\hat{a}^*)^{-1})(1 + \tfrac{1}{12}\lambda^2L^2 + \tfrac{1}{2}L^2\lambda/\hat{a})\right)^{1/2}. \tag{39}$$

The non-asymptotic estimate for the first term in (38) is provided in the following lemma.

Lemma 11. *Let Assumptions 1, 2, 3, 5 and 6 hold. For any* $0 < \lambda < \bar{\lambda}_{\max}$ *given in* (10)*, one obtains*

$$W_2(\mathcal{L}(\dot{\theta}_n^\lambda), \mathcal{L}(\theta_n^\lambda)) \leq \bar{C}_{8,2}\sqrt{\lambda},$$

where

$$\begin{aligned}\bar{C}_{8,2} &= \sqrt{c_5/2\hat{a}^*}\\ c_5 &= (8L^2 + 16L_1^2\mathbb{E}\left[K_\rho(X_0)\right])(\mathbb{E}\left[|\theta_0|^2\right] + \hat{a}^{-1}\bar{c}_4)\\ &\quad + (8L^2 + 40L_1^2\mathbb{E}\left[K_\rho(X_0)\right])|\theta^*|^2 + 24L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + 24\mathbb{E}\left[F_*^2(X_0)\right].\end{aligned} \tag{40}$$

Proof. By using synchronous coupling for the algorithms (34) and (2), one obtains

$$\begin{aligned}|\dot{\theta}_{n+1}^\lambda - \theta_{n+1}^\lambda|^2 &= |\dot{\theta}_n^\lambda - \theta_n^\lambda - \lambda(h(\dot{\theta}_n^\lambda) - H(\theta_n^\lambda, X_{n+1}))|^2\\ &= |\dot{\theta}_n^\lambda - \theta_n^\lambda|^2 - 2\lambda\langle\dot{\theta}_n^\lambda - \theta_n^\lambda, h(\dot{\theta}_n^\lambda) - H(\theta_n^\lambda, X_{n+1})\rangle + \lambda^2|h(\dot{\theta}_n^\lambda) - H(\theta_n^\lambda, X_{n+1})|^2\\ &\leq |\dot{\theta}_n^\lambda - \theta_n^\lambda|^2 - 2\lambda\langle\dot{\theta}_n^\lambda - \theta_n^\lambda, h(\dot{\theta}_n^\lambda) - h(\theta_n^\lambda)\rangle - 2\lambda\langle\dot{\theta}_n^\lambda - \theta_n^\lambda, h(\theta_n^\lambda) - H(\theta_n^\lambda, X_{n+1})\rangle\\ &\quad + 2\lambda^2|h(\dot{\theta}_n^\lambda) - h(\theta_n^\lambda)|^2 + 2\lambda^2|h(\theta_n^\lambda) - H(\theta_n^\lambda, X_{n+1})|^2,\end{aligned}$$

which implies, by taking conditional expectation on both sides and by using Remark 6

$$\begin{aligned}\mathbb{E}\left[|\dot{\theta}_{n+1}^\lambda - \theta_{n+1}^\lambda|^2\middle|\dot{\theta}_n^\lambda, \theta_n^\lambda\right] &\leq |\dot{\theta}_n^\lambda - \theta_n^\lambda|^2 - 2\hat{a}^*\lambda|\dot{\theta}_n^\lambda - \theta_n^\lambda|^2 - \frac{2\lambda}{\hat{a}+L}|h(\dot{\theta}_n^\lambda) - h(\theta_n^\lambda)|^2\\ &\quad + 2\lambda^2|h(\dot{\theta}_n^\lambda) - h(\theta_n^\lambda)|^2 + 2\lambda^2\mathbb{E}\left[|h(\theta_n^\lambda) - H(\theta_n^\lambda, X_{n+1})|^2\middle|\dot{\theta}_n^\lambda, \theta_n^\lambda\right],\end{aligned}$$

where $\hat{a}^* = \hat{a}L/(\hat{a}+L)$. For $\lambda < \bar{\lambda}_{\max}$, one obtains by using Remark 1 and 2

$$\begin{aligned}&\mathbb{E}\left[|\dot{\theta}_{n+1}^\lambda - \theta_{n+1}^\lambda|^2\middle|\dot{\theta}_n^\lambda, \theta_n^\lambda\right]\\ &\leq (1 - 2\hat{a}^*\lambda)|\dot{\theta}_n^\lambda - \theta_n^\lambda|^2 + 4\lambda^2\mathbb{E}\left[|h(\theta_n^\lambda)|^2\middle|\dot{\theta}_n^\lambda, \theta_n^\lambda\right]\\ &\quad + 4\lambda^2\mathbb{E}\left[|H(\theta_n^\lambda, X_{n+1})|^2\middle|\dot{\theta}_n^\lambda, \theta_n^\lambda\right]\\ &\leq (1 - 2\hat{a}^*\lambda)|\dot{\theta}_n^\lambda - \theta_n^\lambda|^2 + 4\lambda^2L^2\mathbb{E}\left[|\theta_n^\lambda - \theta^*|^2\middle|\dot{\theta}_n^\lambda, \theta_n^\lambda\right]\\ &\quad + 4\lambda^2\mathbb{E}\left[\left((1 + |X_{n+1}|)^{\rho+1}(L_1|\theta_n^\lambda - \theta^*| + L_1|\theta^*| + L_2) + F_*(X_{n+1})\right)^2\middle|\dot{\theta}_n^\lambda, \theta_n^\lambda\right]\end{aligned}$$

$$\leq (1-2\hat{a}^*\lambda)|\dot{\theta}_n^\lambda - \theta_n^\lambda|^2 + (4\lambda^2 L^2 + 8\lambda^2 L_1^2 \mathbb{E}\left[K_\rho(X_0)\right])|\theta_n^\lambda - \theta^*|^2$$
$$+ 24\lambda^2 L_1^2 \mathbb{E}\left[K_\rho(X_0)\right]|\theta^*|^2 + 24\lambda^2 L_2^2 \mathbb{E}\left[K_\rho(X_0)\right] + 24\lambda^2 \mathbb{E}\left[F_*^2(X_0)\right].$$

Finally, one calculates by using Lemma 10,

$$\begin{aligned}\mathbb{E}\left[|\dot{\theta}_{n+1}^\lambda - \theta_{n+1}^\lambda|^2\right] &\leq (1-2\hat{a}^*\lambda)\mathbb{E}\left[|\dot{\theta}_n^\lambda - \theta_n^\lambda|^2\right] + (4\lambda^2 L^2 + 8\lambda^2 L_1^2 \mathbb{E}\left[K_\rho(X_0)\right])\mathbb{E}\left[|\theta_n^\lambda - \theta^*|^2\right] \\ &\quad + 24\lambda^2 L_1^2 \mathbb{E}\left[K_\rho(X_0)\right]|\theta^*|^2 + 24\lambda^2 L_2^2 \mathbb{E}\left[K_\rho(X_0)\right] + 24\lambda^2 \mathbb{E}\left[F_*^2(X_0)\right] \\ &\leq (1-2\hat{a}^*\lambda)\mathbb{E}\left[|\dot{\theta}_n^\lambda - \theta_n^\lambda|^2\right] + \lambda^2 c_5,\end{aligned}$$

where $c_5 = (8L^2 + 16L_1^2\mathbb{E}\left[K_\rho(X_0)\right])(\mathbb{E}\left[|\theta_0|^2\right] + \hat{a}^{-1}\bar{c}_4) + (8L^2 + 40L_1^2\mathbb{E}\left[K_\rho(X_0)\right])|\theta^*|^2 + 24L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + 24\mathbb{E}\left[F_*^2(X_0)\right]$. The result follows by induction. $\square$

Proof of Theorem 2 One observes that by using Theorem 3 and Lemma 11

$$\begin{aligned}W_2(\mathcal{L}(\theta_n^\lambda), \pi_\beta) &\leq W_2(\mathcal{L}(\theta_n^\lambda), \mathcal{L}(\dot{\theta}_n^\lambda)) + W_2(\mathcal{L}(\dot{\theta}_n^\lambda), \pi_\lambda) + W_2(\pi_\lambda, \pi_\beta) \\ &\leq \bar{C}_{8,2}\sqrt{\lambda} + \bar{C}_7 e^{-\hat{a}^*\lambda n} + \bar{C}_{8,1}\sqrt{\lambda} \\ &\leq C_7 e^{-C_6\lambda n} + C_8\sqrt{\lambda},\end{aligned}$$

where

$$C_6 = \hat{a}^*, \quad C_7 = \bar{C}_7, \quad C_8 = \bar{C}_{8,1} + \bar{C}_{8,2} \tag{41}$$

with $\hat{a}^* = \hat{a}L/(\hat{a}+L)$, $\bar{C}_7$ given in Lemma 3, $\bar{C}_{8,1}$ and $\bar{C}_{8,2}$ given in (39) and (40) respectively.

Proof of Corollary 4 The proof follows the same lines as the proof of Corollary 2. To obtain an upper bound for the expected excess risk $\mathbb{E}[U(\hat{\theta})] - \inf_{\theta\in\mathbb{R}^d} U(\theta)$, one considers

$$\mathbb{E}[U(\hat{\theta})] - \inf_{\theta\in\mathbb{R}^d} U(\theta) = \Big(\mathbb{E}[U(\hat{\theta})] - \mathbb{E}[U(Z_\infty)]\Big) + \left(\mathbb{E}[U(Z_\infty)] - \inf_{\theta\in\mathbb{R}^d} U(\theta)\right), \tag{42}$$

where $\hat{\theta} = \theta_n^\lambda$ and $Z_\infty \sim \pi_\beta$ with $\pi_\beta(\theta) = \exp(-\beta U(\theta))$ for all $\theta \in \mathbb{R}^d$. By using [19, Lemma 3.5], Lemma 8, 10 and Theorem 2, the first term on the RHS of (42) can be bounded by

$$\begin{aligned}&\mathbb{E}[U(\hat{\theta})] - \mathbb{E}[U(Z_\infty)] \\ &\leq \left(L(\mathbb{E}\left[|\theta_0 - \theta^*|^2\right] + \bar{c}_4\hat{a}^{-1} + |\theta^*|^2)^{1/2} + |h(0)|\right) W_2(\mathcal{L}(\theta_n^\lambda), \pi_\beta) \\ &\leq \left(L(\mathbb{E}\left[|\theta_0 - \theta^*|^2\right] + \bar{c}_4\hat{a}^{-1} + |\theta^*|^2)^{1/2} + |h(0)|\right)\left(C_7 e^{-C_6\lambda n} + C_8\sqrt{\lambda}\right) \\ &\leq \hat{C}_5 e^{-\hat{C}_4\lambda n} + \hat{C}_6\sqrt{\lambda},\end{aligned}$$

where $\theta^* \in \mathbb{R}^d$ is the minimizer of U, and

$$\begin{aligned}\hat{C}_4 &= C_6, \\ \hat{C}_5 &= C_7\left(L(\mathbb{E}\left[|\theta_0 - \theta^*|^2\right] + \bar{c}_4\hat{a}^{-1} + |\theta^*|^2)^{1/2} + |h(0)|\right), \\ \hat{C}_6 &= C_8\left(L(\mathbb{E}\left[|\theta_0 - \theta^*|^2\right] + \bar{c}_4\hat{a}^{-1} + |\theta^*|^2)^{1/2} + |h(0)|\right),\end{aligned} \tag{43}$$

with C_6, C_7, C_8 given in (41) and $\bar{c}_4$ given in (35). Moreover, the second term on the RHS of (42) can be estimated by using [19, Proposition 3.4], which gives,

$$\mathbb{E}[U(Z_\infty)] - \inf_{\theta\in\mathbb{R}^d} U(\theta) \leq \frac{\hat{C}_7}{\beta},$$

where

$$\hat{C}_7 = \frac{d}{2}\log\left(\frac{e\beta L}{d}\left(\frac{d}{\hat{a}\beta} + |\theta^*|^2\right)\right). \tag{44}$$

Finally, one obtains

$$\mathbb{E}[U(\hat{\theta})] - \inf_{\theta\in\mathbb{R}^d} U(\theta) \leq \hat{C}_5 e^{-\hat{C}_4\lambda n} + \hat{C}_6\sqrt{\lambda} + \hat{C}_7/\beta.$$

5 Applications

5.1 Quantile estimation with L_2 regularization

We consider the problem of quantile estimation for AR(1) processes, which has been discussed in [7], [17] and [22] amongst others, with L_2 regularization. It assumed therefore that the data $X_t \in \mathbb{R}$, $t \in \mathbb{Z}$, follows an AR(1) process given by

$$X_{t+1} = \alpha X_t + \bar{\xi}_{t+1},$$

where α is a constant with $|\alpha| < 1$ and $(\bar{\xi}_t)_{t\in\mathbb{Z}}$ are i.i.d. standard Normal random variables. The above expression can be further rewritten as

$$X_t = \sum_{j=0}^{\infty} \alpha^j \bar{\xi}_{t-j}.$$

One notes that X_t has a stationary distribution π_X which is normally distributed with mean 0 and variance $1/(1-\alpha^2)$. Our task is to identify the q-th quantile of the stationary distribution π_X using the SGLD algorithm (2), in other words, we aim to solve the following problem:

$$\min_\theta \mathbb{E}\left[l_q(X_\infty - \theta)\right] + \gamma|\theta|^2,$$

where $X_\infty \sim \pi_X$ and

$$l_q(z) = \begin{cases} qz, & z \geq 0, \\ (q-1)z, & z < 0. \end{cases}$$

The stochastic gradient $H : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is given by

$$H(\theta, x) = -q + \mathbb{1}_{\{x<\theta\}} + 2\gamma\theta, \tag{45}$$

where γ is a positive constant. To check Assumption 1, denote by $F(\theta, x) = -q+2\gamma\theta$, $G(\theta, x) = \mathbb{1}_{\{x<\theta\}}$. It can be easily seen that Assumption 1 holds with $\rho = 0$, $L_1 = 2\gamma$, $L_2 = 0$ and $K_1(x) = 1$. Then, by Remark 3 and its proof in A.1, Assumption 3 holds with $L = 2\gamma + 1$. Moreover, Assumption 4 holds with $A(x) = \gamma \mathbf{I}_d$ and $b(x) = q^2/(4\gamma)$, which implies $a = \gamma$ and $b = q^2/(4\gamma)$.

One notes that the value of the q-th quantile of π_X is given by $\theta^* = N(q)/\sqrt{1-\alpha^2}$ where $N(\cdot)$ is the cumulative distribution function of the standard normal distribution. For the simulation, set $\alpha = 0.5$, $q = 0.95$, and thus, $\theta^* = 1.89$. Moreover, let $m = 1$, $\theta_0 = 3$, $\beta = 10^8$ and $\gamma = 10^{-6}$. Note that we use the step restriction given in Remark 7 for all the examples in this section. In Figure 5.1, the left graph is obtained by using the SGLD algorithm (2) with $\lambda = 10^{-4}$ and the number of iterations $n = 10^6$. It shows the path of θ_n with the first 10000 iterations being discarded, and the path stabilises at around the true value $\theta^* = 1.89$. The right graph of Figure 5.1 illustrates the rate of convergence of the SGLD algorithm in Wasserstein-1 distance based on 5000 samples. The slope of the results in W_1 obtained using numerical experiments is 0.5022, which supports our theoretical finding in Theorem 1 with rate 1/2.

5.2 VaR-CVaR algorithm

In this section, we consider the problem of computing Value-at-Risk (VaR) and Conditional-Value-at-Risk (CVaR), which are two commonly used risk measures in financial risk management. In order to obtain the two quantities, one considers the following optimization problem:

$$\min_\theta V(\theta) = \min_\theta \left(\mathbb{E}\left[\theta + \frac{1}{1-\bar{q}}(f(X) - \theta)_+\right] + \gamma|\theta|^2 \right), \tag{46}$$

where $0 < \bar{q} < 1$, f is continuous and $f(X)$ is integrable with respect to the probability measure. As noted in [4], f can represent more complicated payoff structures than simple vanilla instruments

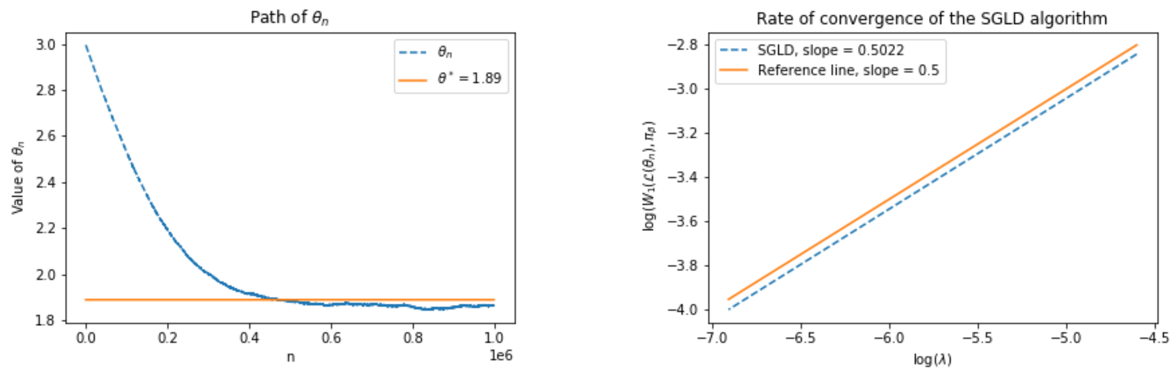

Figure 5.1: *[Left] Path of θ_n when $q = 0.95$. [Right] Rate of convergence of the SGLD algorithm.*

	$\bar{q} = 0.95$				$\bar{q} = 0.99$			
	VaR*	CVaR*	VaR_{SGLD}	$\text{CVaR}_{\text{SGLD}}$	VaR*	CVaR*	VaR_{SGLD}	$\text{CVaR}_{\text{SGLD}}$
$\mu = 0, \sigma = 1$	1.645	2.062	1.642 (0.02)	2.062 (0.0006)	2.326	2.677	2.329 (0.04)	2.662 (0.0038)
$\mu = 1, \sigma = 2$	4.290	5.124	4.294 (0.03)	5.126 (0.0006)	5.653	6.335	5.640 (0.06)	6.336 (0.0032)
$\mu = 3, \sigma = 5$	11.224	13.311	11.230 (0.05)	13.305 (0.0006)	14.632	16.337	14.643 (0.11)	16.313 (0.006)

Table 1: VaR and CVaR for normal distribution $N(\mu, \sigma)$.

	$\bar{q} = 0.95$				$\bar{q} = 0.99$			
	VaR*	CVaR*	VaR_{SGLD}	$\text{CVaR}_{\text{SGLD}}$	VaR*	CVaR*	VaR_{SGLD}	$\text{CVaR}_{\text{SGLD}}$
d.f. = 10	1.812	2.416	1.808 (0.02)	2.407 (0.0005)	2.764	3.357	2.767 (0.05)	3.350 (0.003)
d.f. = 7	1.895	2.595	1.895 (0.03)	2.594 (0.0008)	2.998	3.757	3.001 (0.05)	3.782 (0.0024)
d.f. = 3	2.353	3.876	2.358 (0.03)	3.873 (0.0008)	4.541	6.968	4.542 (0.08)	6.967 (0.0028)

Table 2: VaR and CVaR for Student's t distribution.

while X can accommodate a large family of asset distributions including those generated by stochastic/local volatility models, see e.g. [11], [20] and [21] references therein. Then, by [4, Proposition 2.1], $\text{VaR}_{\bar{q}}(f(X)) = \operatorname{argmin} V(\theta)$ and $\text{CVaR}_{\bar{q}}(f(X)) = \min_\theta V(\theta)$. To compute VaR, the stochastic gradient $H : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ of the SGLD algorithm (2) is given by

$$H(\theta, x) = 1 - \frac{1}{1-\bar{q}} \mathbb{1}_{\{f(x) \geq \theta\}} + 2\gamma\theta = -\frac{\bar{q}}{1-\bar{q}} + \frac{1}{1-\bar{q}} \mathbb{1}_{\{f(x) < \theta\}} + 2\gamma\theta.$$

5.2.1 Single asset

Let $f(x) = x$, one notices that the above expression has a similar form as (45). Then, one can check that Assumption 1 - 4 are satisfied. More precisely, denote by $F(\theta, x) = -\bar{q}/(1-\bar{q}) + 2\gamma\theta$, $G(\theta, x) = \mathbb{1}_{\{x<\theta\}}/(1-\bar{q})$, Assumption 1 holds with $\rho = 0$, $L_1 = 2\gamma$, $L_2 = 0$ and $K_1(x) = 1/(1-\bar{q})$. Let X be a one-dimensional random variable with finite fourth moment, then Assumption 2 is satisfied. Denote by $\bar{c}_d$ the upper bound of the density of X, Assumption 3 holds with $L = 2\gamma + \bar{c}_d/(1-\bar{q})$. Furthermore, Assumption 4 holds with $A(x) = \gamma \mathbf{I}_d$ and $b(x) = \bar{q}^2/(4\gamma(1-\bar{q})^2)$, which implies $a = \gamma$ and $b = \bar{q}^2/(4\gamma(1-\bar{q})^2)$.

For the numerical experiments, we set $\theta_0 = 0$, $\beta = 10^8$, $\gamma = 10^{-8}$, $\lambda = 10^{-4}$ and the number of iterations $n = 10^6$. Table 1 and 2 present VaR and CVaR for the normal distribution and Student's t-distribution. VaR* and CVaR* in the tables denote the theoretical values, while VaR_{SGLD} and $\text{CVaR}_{\text{SGLD}}$ denote the numerical approximations from the SGLD algorithm (2). Each approximation in the table is obtained based on 10000 samples, which is followed by its sample standard deviation

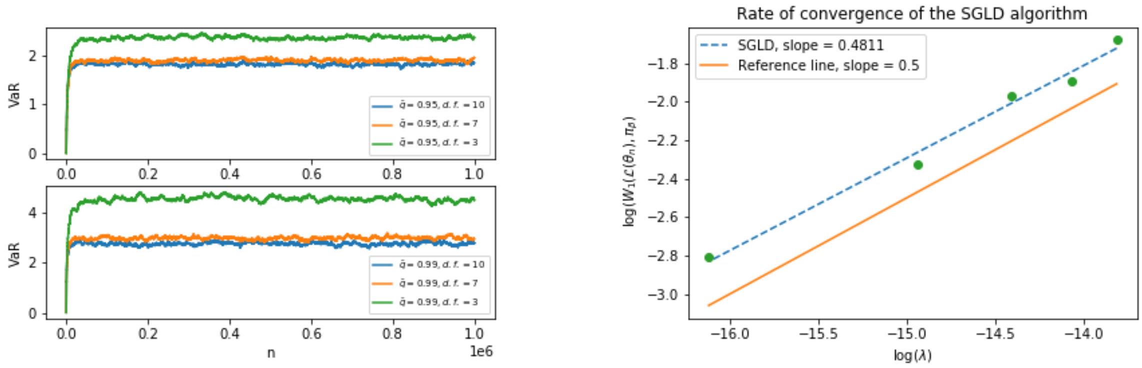

Figure 5.2: *[Left] Path of θ_n (VaR) for Student's t-distribution. [Right] Rate of convergence of the SGLD algorithm based on 5000 samples.*

shown in brackets. In addition, in Figure 5.2, the left graph illustrates the path of θ_n for the t-distribution, whereas the right graph shows that the rate of convergence of the SGLD algorithm (2) is 0.4811. One notes that the samples from π_β is generated by running the SGLD algorithm with $\lambda = 10^{-5}$.

5.2.2 Minimizing CVaR of portfolios of assets

To minimize CVaR for a given portfolio, we consider the following optimization problem:

$$\min_{\hat{\theta}} V(\hat{\theta}) = \min_{\hat{\theta}} \left(\mathbb{E}\left[\frac{1}{1-\bar{q}} \left(\sum_{i=1}^{n} g_i(w) X_i - \theta \right)_+ + \theta \right] + \gamma |\hat{\theta}|^2 \right), \tag{47}$$

where the parameter $\hat{\theta} := (\theta, w)^\mathsf{T} = (\theta, w_1, \ldots, w_n)^\mathsf{T}$ and $g_i(w) := \frac{e^{w_i}}{\sum_{j=1}^{n} e^{w_j}} \in (0, 1)$ for $i = 1, \ldots, n$. By solving (47), we obtain not only VaR for a given portfolio, but also the optimal weight for each asset in the portfolio such that CVaR is minimized.

For reasons of brevity, we assume here that the X_i's, for $i = 1, \ldots, n$, are i.i.d. one-dimensional random variables (with finite fourth moments). Our results can be naturally extended to the case of dependent data streams via the concept of L-mixing as explained in [8].

Let c_X, $c_{\bar{X}}$ denote the first and second absolute moment respectively of X_1. Moreover, let $|x| f_{X_i}(x)$ be bounded for any i and $x \in \mathbb{R}$. Note that this latter requirement is satisfied for a wide range of distributions, for example, the distributions shown in Table 3. Then, the stochastic gradient $H_{\hat{\theta}}(\hat{\theta}, x) : \mathbb{R}^{n+1} \times \mathbb{R}^n \to \mathbb{R}^{n+1}$ is defined as

$$H_{\hat{\theta}}(\hat{\theta}, x) := (H_\theta(\hat{\theta}, x), H_{w_1}(\hat{\theta}, x), \ldots, H_{w_n}(\hat{\theta}, x))^\mathsf{T},$$

where $H_\theta(\hat{\theta}, x) : \mathbb{R}^{n+1} \times \mathbb{R}^n \to \mathbb{R}$ and $H_{w_j}(\hat{\theta}, x) : \mathbb{R}^{n+1} \times \mathbb{R}^n \to \mathbb{R}$ for all j are given by

$$H_\theta(\hat{\theta}, x) = 1 - \frac{1}{1-\bar{q}} \mathbb{1}_{\{\sum_{i=1}^{n} g_i(w) x_i \geq \theta\}} + 2\gamma\theta,$$

and

$$H_{w_j}(\hat{\theta}, x) = \frac{1}{1-\bar{q}} \hat{g}_{w_j}(w, x) \mathbb{1}_{\{\sum_{i=1}^{n} g_i(w) x_i \geq \theta\}} + 2\gamma w_j,$$

where

$$\hat{g}_{w_j}(w, x) = \sum_{i=1}^{n} \frac{\partial g_i(w)}{\partial w_j} x_i$$

for any $j = 1, \ldots, n$ with $\frac{\partial g_j(w)}{\partial w_j} = \frac{e^{w_j}(\sum_{l \neq j} e^{w_l})}{(\sum_{l=1}^{n} e^{w_l})^2}$, and $\frac{\partial g_i(w)}{\partial w_j} = -\frac{e^{w_i} e^{w_j}}{(\sum_{l=1}^{n} e^{w_l})^2}$ for $i \neq j$. One notes that $|\hat{g}_{w_j}(w, x)| \leq \sum_{i=1}^{n} |x_i|$ for any j. Moreover, if Assumption 1 - 4 hold for H_θ and H_{w_j} for any j, then the assumptions hold for $H_{\hat{\theta}}$.

We first check assumptions for H_θ. Denote by

$$F_\theta(\hat\theta, x) = 2\gamma\theta, \quad G_\theta(\hat\theta, x) = 1 - \mathbb{1}_{\{\sum_{i=1}^n g_i(w)x_i \geq \theta\}}/(1-\bar q),$$

then $H_\theta = F_\theta + G_\theta$. Assumption 1 holds with $\rho = 0$, $L_1 = 2\gamma$, $L_2 = 0$ and $K_1(x) = (2-\bar q)/(1-\bar q)$. By taking into consideration the expression of $K_1(x)$ and the construction of the problem, Assumption 2 is satisfied. Assumption 4 holds with $A(x) = 2\gamma \mathbf{I}_d$ and $b(x) = 0$, which implies $a = 2\gamma$ and $b = 0$. To check Assumption 3, one considers $\hat\theta' := (\bar\theta, w)^\intercal$, and then calculates by assuming without loss of generality $g_n(w) = \max\{g_1(w), \dots, g_n(w)\}$

$$\begin{aligned}
&\mathbb{E}\left[\left|H_\theta(\hat\theta, X) - H_\theta(\hat\theta', X)\right|\right] \\
&\leq 2\gamma\left|\theta - \bar\theta\right| + \frac{1}{1-\bar q}\mathbb{E}\left[\left|\mathbb{1}_{\{\sum_{i=1}^n g_i(w)X_i \geq \theta\}} - \mathbb{1}_{\{\sum_{i=1}^n g_i(w)X_i \geq \bar\theta\}}\right|\right] \\
&\leq 2\gamma\left|\hat\theta - \hat\theta'\right| + \frac{1}{1-\bar q}(E_1 + E_2),
\end{aligned}$$

where

$$E_1 = \mathbb{E}\left[\mathbb{1}_{\{\theta \leq \sum_{i=1}^n g_i(w)X_i \leq \bar\theta\}}\right], \quad E_2 = \mathbb{E}\left[\mathbb{1}_{\{\bar\theta \leq \sum_{i=1}^n g_i(w)X_i \leq \theta\}}\right].$$

To estimate E_1, one writes

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{\{\theta \leq \sum_{i=1}^n g_i(w)X_i \leq \bar\theta\}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\{(\theta - \sum_{i\neq n} g_i(w)X_i)/g_n(w)) \leq X_n \leq (\bar\theta - \sum_{i \neq n} g_i(w)X_i)/g_n(w))\}} \,\middle|\, X_1, \dots, X_{n-1}\right]\right] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{(\theta - \sum_{i\neq n} g_i(w)x_i)/g_n(w)}^{(\bar\theta - \sum_{i\neq n} g_i(w)x_i)/g_n(w)} f_{X_n}(z)\, dz f_{X_{n-1}}(x_{n-1})\, dx_{n-1} \cdots f_{X_1}(x_1)\, dx_1 \\
&\leq n c_{X_n}\left|\hat\theta - \hat\theta'\right|,
\end{aligned}$$

where we use the fact $g_n(w) \geq 1/n$ in the last inequality and c_{X_n} denotes the upper bound of the density of X_n. E_2 can be estimated by using similar arguments. Then, one obtains

$$\mathbb{E}\left[\left|H_\theta(\hat\theta, X) - H_\theta(\hat\theta', X)\right|\right] \leq (2\gamma + 2nc_{X_n}/(1-\bar q))\left|\hat\theta - \hat\theta'\right|,$$

which implies Assumption 3 holds with $L = 2\gamma + 2nc_{X_n}/(1-\bar q)$.

Next, we check assumptions for H_{w_j}. Denote by

$$F_{w_j}(\hat\theta, x) = 2\gamma w_j, \quad G_{w_j}(\hat\theta, x) = \hat g_{w_j}(w, x)\mathbb{1}_{\{\sum_{i=1}^n g_i(w)x_i \geq \theta\}}/(1-\bar q),$$

then $H_{w_j} = F_{w_j} + G_{w_j}$. Assumption 1 holds with $\rho = 0$, $L_1 = 2\gamma$, $L_2 = 0$ and $K_1(x) = \sum_i |x_i|/(1-\bar q)$. By taking into consideration the expression of $K_1(x)$ and the construction of the problem, Assumption 2 is satisfied. Assumption 4 holds with $A(x) = 2\gamma\mathbf{I}_d$ and $b(x) = 0$, which implies $a = 2\gamma$ and $b = 0$. Then, we check Assumption 3 for H_{w_1}, and the arguments stay the same lines for any other H_{w_j}, $j = 2, \dots, n$. Consider $\hat\theta^\sharp := (\theta, \bar w)^\intercal = (\theta, \bar w_1, w_2, \dots, w_n)^\intercal$. Then, one calculates

$$\begin{aligned}
&\mathbb{E}\left[\left|H_{w_1}(\hat\theta, X) - H_{w_1}(\hat\theta^\sharp, X)\right|\right] \\
&\leq 2\gamma|w_1 - \bar w_1| + \frac{1}{1-\bar q}\mathbb{E}\left[\left|\hat g_{w_1}(w, X)\mathbb{1}_{\{\sum_{i=1}^n g_i(w)X_i \geq \theta\}} - \hat g_{w_1}(\bar w, X)\mathbb{1}_{\{\sum_{i=1} g_i(\bar w)X_i \geq \theta\}}\right|\right] \\
&\leq 2\gamma\left|\hat\theta - \hat\theta^\sharp\right| + \frac{1}{1-\bar q}\mathbb{E}\left[\left|\hat g_{w_1}(w, X)\mathbb{1}_{\{\sum_{i=1}^n g_i(w)X_i \geq \theta\}} - \hat g_{w_1}(\bar w, X)\mathbb{1}_{\{\sum_{i=1}^n g_i(w)X_i \geq \theta\}}\right|\right] \\
&\quad + \frac{1}{1-\bar q}\mathbb{E}\left[\left|\hat g_{w_1}(\bar w, X)\mathbb{1}_{\{\sum_{i=1}^n g_i(w)X_i \geq \theta\}} - \hat g_{w_1}(\bar w, X)\mathbb{1}_{\{\sum_{i=1} g_i(\bar w)X_i \geq \theta\}}\right|\right] \\
&\leq 2\gamma\left|\hat\theta - \hat\theta^\sharp\right| + \frac{2nc_X}{1-\bar q}|w_1 - \bar w_1|
\end{aligned}$$

		SGLD algorithm				Reference			
X_1	X_2	w_1	w_2	$g_1(w)X_1 + g_2(w)X_2$		w_1^*	w_2^*	$g_1(w^*)X_1 + g_2(w^*)X_2$	
				$\mathrm{VaR_{SGLD}}$	$\mathrm{CVaR_{SGLD}}$			VaR*	CVaR*
$N(500,1)$	$N(0,10^{-4})$	0.00002	0.99998	0.025	0.03	0	1	0.016	0.021
$N(0,10^6)$	$N(0,10^{-4})$	0.000006	0.999994	0.016	0.25	0	1	0.016	0.021
$N(1,4)$	$N(0,1)$	0.111	0.889	1.615	2.004	0.11	0.89	1.617	1.999
$N(0,1)$	t with d.f. $= 2.01$	0.917	0.083	1.567	1.975	0.9	0.1	1.531	1.971
$N(0,1)$	t with d.f. $= 10$	0.577	0.423	1.236	1.554	0.58	0.42	1.224	1.553
$N(0,1)$	t with d.f. $= 1000$	0.503	0.497	1.15	1.46	0.5	0.5	1.165	1.461
$N(1,4)$	t with d.f. $= 2.01$	0.596	0.404	2.941	4.130	0.61	0.39	2.985	4.115
$N(1,4)$	t with d.f. $= 10$	0.172	0.828	1.743	2.290	0.17	0.83	1.779	2.286
$N(1,4)$	t with d.f. $= 1000$	0.113	0.887	1.594	2.008	0.11	0.89	1.619	2.002
$N(0,1)$	Logistic(0,1)	0.775	0.225	1.422	1.816	0.78	0.22	1.442	1.813
$N(0,1)$	Logistic(0,29)	0.999	0.001	1.633	2.110	1	0	1.645	2.063
$N(0,1)$	Logistic(2,10)	0.997	0.003	1.650	2.101	1	0	1.648	2.065
$N(1,4)$	Logistic(0,1)	0.402	0.598	2.635	3.262	0.4	0.6	2.607	3.261
$N(1,4)$	Logistic(0,29)	0.998	0.002	4.284	5.145	1	0	4.284	5.116
$N(1,4)$	Logistic(2,10)	0.991	0.009	4.255	5.132	0.99	0.01	4.283	5.114
$N(0,1)$	Lognormal(0,1)	0.966	0.034	1.662	2.068	0.97	0.03	1.647	2.054
$N(0,1)$	Lognormal(0,0.01)	0.074	0.926	1.145	1.205	0.07	0.93	1.132	1.186
$N(0,1)$	Lognormal(1,4)	0.9997	0.0003	1.674	2.136	1	0	1.645	2.062
$N(1,4)$	Lognormal(0,1)	0.732	0.268	3.750	4.6050.74	0.74	0.26	3.771	4.599
$N(1,4)$	Lognormal(0,0.01)	0.010	0.0.989	1.173	1.301	0	1	1.179	1.230
$N(1,4)$	Lognormal(1,4)	0.997	0.003	4.266	5.194	1	0	4.292	5.129
Logistic(0,1)	Lognormal(0,1)	0.817	0.183	2.797	3.727	0.81	0.19	2.814	3.724
Logistic(0,1)	Lognormal(0,0.01)	0.022	0.978	1.169	1.256	0.02	0.98	1.164	1.217
Logistic(0,1)	Lognormal(1,4)	0.997	0.003	2.961	4.030	1	0	2.947	3.971
Logistic(2,10)	Lognormal(0,1)	0.043	0.956	5.245	8.412	0.04	0.96	5.198	8.400
Logistic(2,10)	Lognormal(0,0.01)	0.009	0.991	1.184	1.315	0	1	1.179	1.229
Logistic(2,10)	Lognormal(1,4)	0.996	0.004	31.651	41.748	0.99	0.01	31.420	41.738

Table 3: 95% VaR and CVaR for portfolios of two assets X_1, X_2 with the form $w_1X_1 + w_2X_2$.

$$
\begin{aligned}
&+ \frac{1}{1-\bar{q}}\mathbb{E}\left[\left|\hat{g}_{w_1}(\bar{w},X)\mathbb{1}_{\{\sum_{i=1}^n g_i(w)X_i\geq\theta\}} - \hat{g}_{w_1}(\bar{w},X)\mathbb{1}_{\{\sum_{i=1} g_i(\bar{w})X_i\geq\theta\}}\right|\right] \\
&\leq 2\gamma\left|\hat{\theta}-\hat{\theta}^\sharp\right| + \frac{2nc_X}{1-\bar{q}}\left|\hat{\theta}-\hat{\theta}^\sharp\right| \\
&+ \frac{1}{1-\bar{q}}\mathbb{E}\left[\left|\hat{g}_{w_1}(\bar{w},X)\mathbb{1}_{\{\sum_{i=1}^n g_i(w)X_i\geq\theta\}} - \hat{g}_{w_1}(\bar{w},X)\mathbb{1}_{\{\sum_{i=1} g_i(\bar{w})X_i\geq\theta\}}\right|\right],
\end{aligned}
$$

where the third inequality holds due to the fact that $|\hat{g}_{w_1}(w,X) - \hat{g}_{w_1}(\bar{w},X)| \leq 2|w_1-\bar{w}_1|\sum_i |X_i|$. Then, by using $|\hat{g}_{w_1}(\bar{w},x)| \leq \sum_i |x_i|$,

$$
\begin{aligned}
&\mathbb{E}\left[\left|H_{w_1}(\hat{\theta},X) - H_{w_1}(\hat{\theta}^\sharp,X)\right|\right] \\
&\leq 2\gamma\left|\hat{\theta}-\hat{\theta}^\sharp\right| + \frac{2nc_X}{1-\bar{q}}\left|\hat{\theta}-\hat{\theta}^\sharp\right| \\
&\quad + \frac{1}{1-\bar{q}}\mathbb{E}\left[\sum_i |X_i|\left|\mathbb{1}_{\{\sum_{i=1}^n g_i(w)X_i\geq\theta\}} - \mathbb{1}_{\{\sum_{i=1} g_i(\bar{w})X_i\geq\theta\}}\right|\right] \\
&\leq (2\gamma + 2nc_X/(1-\bar{q}))\left|\hat{\theta}-\hat{\theta}^\sharp\right| \\
&\quad + 2(n-1)(c_X(\bar{c}_{X_n}+\bar{c}_{X_1}) + (c_{\bar{X}} + (n-2)c_X^2)(c_{X_n}+c_{X_1}))/(1-\bar{q})\left|\hat{\theta}-\hat{\theta}^\sharp\right|,
\end{aligned} \tag{48}
$$

where c_X, $c_{\bar{X}}$ denote the first and the second absolute moment of X_i's respectively, for any i, $\bar{c}_{X_i}$ is the upper bound of the function $|x|f_{X_i}$, and c_{X_i} is the upper bound of the density of X_i. Detailed calculations to obtain the last inequality in (48) is given in Appendix A.3. Thus Assumption 3 holds with $L = 2\gamma + 2nc_X/(1-\bar{q}) + 2(n-1)(c_X(\bar{c}_{X_n}+\bar{c}_{X_1}) + (c_{\bar{X}} + (n-2)c_X^2)(c_{X_n}+c_{X_1}))/(1-\bar{q})$.

For the numerical experiments, we set $\theta_0 = 0$, $\beta = 10^8$, $\gamma = 10^{-8}$, $\lambda = 10^{-4}$ and the number of iterations $n = 10^6$. Tabel 3 illustrates 95% VaR and CVaR obtained using the SGLD algorithm for

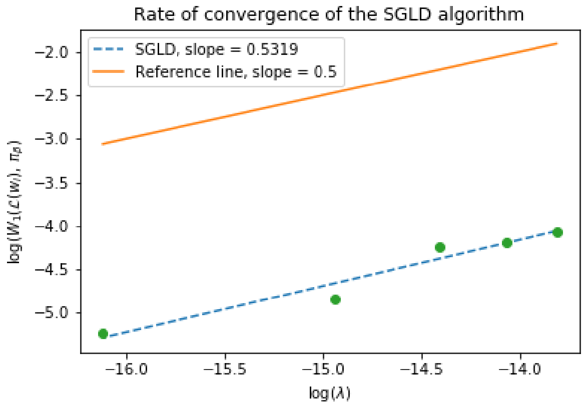

Figure 5.3: *Rate of convergence of the SGLD algorithm for w_1 based on 5000 samples.*

a portfolio of two assets X_1 and X_2 with weights $g_1(w)$ and $g_2(w)$ respectively. The reference values w_1^*, w_2^*, VaR* and CVaR* are obtained numerically in the following way:

1. First, we create 100 evenly spaced numbers over the interval $[0, 1]$.
2. Then, for any given distributions of X_1 and X_2, assign each of the 100 numbers to $g_1(w)$, which is the weight of X_1, and calculate the 95% CVaR for the combination $g_1(w)X_1 + g_2(w)X_2$.
3. Finally, we obtain the minimum CVaR and the corresponding $g_1(w)$ among the 100 values. We denote them as CVaR* and $g_1(w^*)$. Here , one notes that the corresponding VaR* can be calculated using the optimal weights $g_1(w^*)$ and $g_2(w^*)$.

Figure 5.3 shows that the rate of convergence of the SGLD algorithm (2) for the parameter w_1 is 0.5319, which supports the theoretical finding in Theorem 1. One notes that the samples from π_β is generated by running the SGLD algorithm with $\lambda = 10^{-5}$.

References

[1] M. Barkhagen, N. H. Chau, É. Moulines, M. Rásonyi, S. Sabanis and Y. Zhang. On stochastic gradient Langevin dynamics with stationary data streams in the logconcave case. *Preprint*, 2018. arXiv:1812.02709

[2] N. Brosse, A. Durmus and E. Moulines. The promises and pitfalls of stochastic gradient Langevin dynamics. *Advances in Neural Information Processing Systems*, 8268-8278, 2018.

[3] M. Benam, J.C. Fort and G. Pagès, Convergence of the one-dimensional Kohonen algorithm. *Advances in Applied Probability*, 30(3), 850-869, 1998.

[4] O. Bardou, N. Frikha and G. Pagès, Computing VaR and CVaR using stochastic approximation and adaptive unconstrained importance sampling. *Monte Carlo Methods and Applications*, 15(3), 173-210, 2009.

[5] H. Cardot, P. Cénac and P. A. Zitt. Recursive estimation of the conditional geometric median in Hilbert spaces. *Electronic Journal of Statistics*, 6: 2535-2562, 2012.

[6] H. Cardot, P. Cénac and P. A. Zitt. Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli*, 19(1): 18-43, 2013.

[7] N. H. Chau, Ch. Kumar, M. Rásonyi and S. Sabanis. On fixed gain recursive estimators with discontinuity in the parameters. *ESAIM Probability and Statistics*, 23:217–244, 2019.

[8] N. H. Chau, É. Moulines, M. Rásonyi, S. Sabanis and Y. Zhang. On stochastic gradient Langevin dynamics with dependent data streams: the fully non-convex case. *Preprint*, 2019. arXiv:1905.13142

[9] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *The Annals of Probability*, 32(3B), 2702–2732, 2004.

[10] A. S. Dalalyan and A. Karagulyan. User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *To appear in Stochastic Processes and their Applications*, 2019.

[11] B. Dupire, Pricing with a Smile. *Risk*, 7(1):18–20, 1994.

[12] A. Durmus and É. Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Preprint*, 2018. arXiv:1605.01559v3

[13] A. Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields*, 166:851–886, 2016.

[14] A. Eberle, A. Guillin and R. Zimmer. Quantitative Harris-type theorems for diffusions and McKean-Vlasov processes. *In Press, Transactions of the American Mathematical Society*, 2018. https://doi.org/10.1090/tran/7576

[15] G. Fort, É. Moulines, A. Schreck and M. Vihola. Convergence of Markovian Stochastic Approximation with discontinuous dynamics. *SIAM Journal on Control and Optimization*, 54(2): 866–893, 2016.

[16] C.R. Hwang Laplace's method revisited: weak convergence of probability measures. *The Annals of Probability*, 8(6): 1177-1182, 1980.

[17] R. Koenker and G. Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, 33–50, 1978.

[18] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course. Applied Optimization.* Springer, 2004.

[19] M. Raginsky, A. Rakhlin, and M. Telgarsky. Non-convex learning via Stochastic Gradient Langevin Dynamics: a nonasymptotic analysis. *Proceedings of Machine Learning Research*, (65)1674–1703, 2017.

[20] S. Sabanis. Stochastic volatility and the mean reverting process. *Journal of futures markets*, 23(1):33–47, 2003.

[21] S. Sabanis. Stochastic volatility. *International Journal of Theoretical and Applied Finance*, 5(5):515–530, 2002.

[22] I. Takeuchi, Q. V. Le, T. D. Sears and A.J. Smola. Nonparametric quantile estimation. *Journal of machine learning research*, 7(Jul): 1231–1264, 2006.

[23] P. Xu, J. Chen, D. Zhou and Q. Gu. Global convergence of Langevin dynamics based algorithms for nonconvex optimization. *Advances in Neural Information Processing Systems*, 3122-3133, 2018.

A Appendix

A.1 Proof of the claim in Remark 3

We adapt the proof from [7, Lemma 4.7] and extend it to an $\mathbb{R}^m$-valued random variable X_0. It suffices to consider $H(\theta, X_0) = \dot{g}(\theta, X_0)\mathbb{1}_{\bigcap_{i=1}^m \{X_0^{(i)} \in I_i(\theta)\}}$, where $\theta \in \mathbb{R}^d$, $\dot{g}$ is bounded and jointly Lipschitz continuous, i.e. there exist $L_3, L_4, K_2 > 0$ such that for any $\theta, \theta' \in \mathbb{R}^d$, $x, x' \in \mathbb{R}^m$,

$$|\dot{g}(\theta, x) - \dot{g}(\theta', x')| \leq (1 + |x| + |x'|)^\rho (L_3|\theta - \theta'| + L_4|x - x'|), \quad |\dot{g}(\theta, x)| \leq K_2,$$

and the intervals $I_i(\theta)$ take the form $(-\infty, \bar{g}^{(i)}(\theta))$ with $\bar{g}^{(i)}$ Lipschitz. One notices that the proof follows the same lines when $I_i(\theta)$ takes the form $(\bar{g}^{(i)}(\theta), \infty)$, $(\tilde{g}^{(i)}(\theta), \hat{g}^{(i)}(\theta))$ with $\bar{g}^{(i)}, \tilde{g}^{(i)}, \hat{g}^{(i)}$ Lipschitz. One writes,

$$\begin{aligned}
\left|H(\theta, X_0) - H(\theta', X_0)\right| &\leq \left|\dot{g}(\theta, X_0)\mathbb{1}_{\bigcap_{i=1}^m \left\{X_0^{(i)} < \bar{g}^{(i)}(\theta)\right\}} - \dot{g}(\theta', X_0)\mathbb{1}_{\bigcap_{i=1}^m \left\{X_0^{(i)} < \bar{g}^{(i)}(\theta')\right\}}\right| \\
&\leq \left|\dot{g}(\theta, X_0)\mathbb{1}_{\bigcap_{i=1}^m \left\{X_0^{(i)} < \bar{g}^{(i)}(\theta)\right\}} - \dot{g}(\theta', X_0)\mathbb{1}_{\bigcap_{i=1}^m \left\{X_0^{(i)} < \bar{g}^{(i)}(\theta)\right\}}\right| \\
&\quad + \left|\dot{g}(\theta', X_0)\mathbb{1}_{\bigcap_{i=1}^m \left\{X_0^{(i)} < \bar{g}^{(i)}(\theta)\right\}} - \dot{g}(\theta', X_0)\mathbb{1}_{\bigcap_{i=1}^m \left\{X_0^{(i)} < \bar{g}^{(i)}(\theta')\right\}}\right| \\
&\leq L_3(1 + 2|X_0|)^\rho|\theta - \theta'| + K_2\mathbb{1}_{\bigcap_{i=1}^m \left\{X_0^{(i)} \in [\bar{g}^{(i)}(\theta), \bar{g}^{(i)}(\theta'))\right\}},
\end{aligned}$$

where $K_\rho(x)$ for any $x \in \mathbb{R}^m$ is defined in (5) and we assume without loss of generality $\bar{g}^{(i)}(\theta) \leq \bar{g}^{(i)}(\theta')$ for all $i = 1, \ldots, m$. By taking expectation on both sides and by using Cauchy-Schwarz inequality, one obtains

$$\begin{aligned}
&\mathbb{E}\left[\left|H(\theta, X_0) - H(\theta', X_0)\right|\right] \\
&\leq L_3\mathbb{E}[(1 + 2|X_0|)^\rho]|\theta - \theta'| + K_2 P\left(\bigcap_{i=1}^m \left\{X_0^{(i)} \in [\bar{g}^{(i)}(\theta), \bar{g}^{(i)}(\theta'))\right\}\right) \\
&\leq L_3\mathbb{E}[(1 + 2|X_0|)^\rho]|\theta - \theta'| + K_2 \int_{\bar{g}^{(m)}(\theta)}^{\bar{g}^{(m)}(\theta')} \cdots \int_{\bar{g}^{(1)}(\theta)}^{\bar{g}^{(1)}(\theta')} f_{X_0}(x^{(1)}, \ldots, x^{(m)})dx^{(1)} \cdots dx^{(m)} \\
&\leq L_3\mathbb{E}[(1 + 2|X_0|)^\rho]|\theta - \theta'| + K_2 \int_{\bar{g}^{(1)}(\theta)}^{\bar{g}^{(1)}(\theta')} f_{X_0^{(1)}}(x^{(1)})dx^{(1)} \\
&\leq L_3\mathbb{E}[(1 + 2|X_0|)^\rho]|\theta - \theta'| + K_2K_3L_5|\theta - \theta'| \\
&\leq (L_3 + K_2K_3L_5)\mathbb{E}[(1 + 2|X_0|)^\rho]|\theta - \theta'|,
\end{aligned}$$

where $f_{X_0^{(i)}}$ denotes the marginal density function of $X_0^{(i)}$, K_3 is an upper bound of $f_{X_0^{(1)}}$ and L_5 is a Lipschitz constant for $\bar{g}^{(1)}$. Taking $L = L_3 + K_2K_3L_5$ completes the proof.

A.2 Proof of the claim in Remark 4

By Assumption 5, one obtains, for $\theta \in \mathbb{R}^d$ and $x \in \mathbb{R}^m$,

$$\langle F(\theta, x) - F(0, x), \theta\rangle \geq \langle\theta, \hat{A}_1(x)\theta\rangle,$$

which implies

$$\begin{aligned}
\langle F(\theta, x), \theta\rangle &\geq \langle\theta, \hat{A}_1(x)\theta\rangle + \langle F(0, x), \theta\rangle \\
&\geq \langle\theta, \hat{A}_1(x)\theta\rangle - |F(0, x)||\theta| \\
&\geq \langle\theta, \hat{A}_1(x)\theta\rangle - \epsilon|\theta|^2 - (L_2(1 + |x|)^{\rho+1} + |F(0, 0)|)^2/(4\epsilon) \\
&\geq \langle\theta, \hat{A}_1^*(x)\theta\rangle - \hat{b}(x),
\end{aligned}$$

where the third inequality holds due to Assumption 1 and $ab < \epsilon a^2 + b^2/(4\epsilon)$, for any $a, b > 0$, $\epsilon > 0$, $\hat{A}_1^*(x) = \hat{A}_1(x) - \epsilon\mathbf{I}_d$ and $\hat{b}(x) = (L_2(1 + |x|)^{\rho+1} + |F(0, 0)|)^2/(4\epsilon)$.

A.3 Validity of Assumption 3 for VaR-CVaR algorithm in Section 5.2

We aim to show Assumption 3 is valid for H_{w_1}. To achieve this, it is enough to prove

(1) The inequality $|\hat{g}_{w_1}(w, X) - \hat{g}_{w_1}(\bar{w}, X)| \le 2|w_1 - \bar{w}_1| \sum_i |X_i|$ holds, and

(2) the last inequality in (48) is satisfied.

To prove $|\hat{g}_{w_1}(w, X) - \hat{g}_{w_1}(\bar{w}, X)| \le 2|w_1 - \bar{w}_1| \sum_i |X_i|$, recall that for every $j = 1, \dots, n$, $i \neq j$,

$$\frac{\partial g_j(w)}{\partial w_j} = \frac{e^{w_j}(\sum_{l \neq j} e^{w_l})}{(\sum_{l=1}^n e^{w_l})^2}, \quad \frac{\partial g_i(w)}{\partial w_j} = -\frac{e^{w_i} e^{w_j}}{(\sum_{l=1}^n e^{w_l})^2}.$$

Then, one calculates

$$\begin{aligned}
&|\hat{g}_{w_1}(w, X) - \hat{g}_{w_1}(\bar{w}, X)| \\
&= \left| \sum_{i=1}^n \frac{\partial g_i(w)}{\partial w_1} X_i - \sum_{i=1}^n \frac{\partial g_i(\bar{w})}{\partial w_1} X_i \right| \\
&\le \left| \frac{\partial g_1(w)}{\partial w_1} - \frac{\partial g_1(\bar{w})}{\partial w_1} \right| |X_1| + \left| \sum_{i \neq 1} \frac{\partial g_i(w)}{\partial w_1} X_i - \sum_{i \neq 1} \frac{\partial g_i(\bar{w})}{\partial w_1} X_i \right| \\
&\le \left| \frac{e^{w_1}(\sum_{l \neq 1} e^{w_l})}{(\sum_{l=1}^n e^{w_l})^2} - \frac{e^{\bar{w}_1}(\sum_{l \neq 1} e^{w_l})}{(\sum_{l \neq 1} e^{w_l} + e^{\bar{w}_1})^2} \right| |X_1| + \sum_{i \neq 1} \left| \frac{e^{w_i} e^{\bar{w}_1}}{(\sum_{l \neq 1} e^{w_l} + e^{\bar{w}_1})^2} - \frac{e^{w_i} e^{w_1}}{(\sum_{l=1}^n e^{w_l})^2} \right| |X_i| \\
&= \frac{\sum_{l \neq 1} e^{w_l}}{(\sum_{l=1}^n e^{w_l})^2 (\sum_{l \neq 1} e^{w_l} + e^{\bar{w}_1})^2} \left| \left(\sum_{l \neq 1} e^{w_l} \right)^2 \left(e^{w_1} - e^{\bar{w}_1} \right) + e^{\bar{w}_1} e^{w_1} \left(e^{\bar{w}_1} - e^{w_1} \right) \right| |X_1| \\
&\quad + \sum_{i \neq 1} \frac{e^{w_i}}{(\sum_{l=1}^n e^{w_l})^2 (\sum_{l \neq 1} e^{w_l} + e^{\bar{w}_1})^2} \left| \left(\sum_{l \neq 1} e^{w_l} \right)^2 \left(e^{\bar{w}_1} - e^{w_1} \right) + e^{\bar{w}_1} e^{w_1} \left(e^{w_1} - e^{\bar{w}_1} \right) \right| |X_i| \\
&\le 2|w_1 - \bar{w}_1| \sum_{i=1}^n |X_i|,
\end{aligned}$$

where the last inequality holds due to $1 - e^{-x} \le x$ for all $x \ge 0$.

To prove the last inequality in (48) is satisfied, we assume without loss of generality $g_n(w) = \max\{g_2(w), \dots, g_n(w)\}$. Then,

(i) For $\bar{w}_1 \ge w_1$, one calculates

$$\mathbb{E} \left[\sum_i |X_i| \left| \mathbb{1}_{\{\sum_{i=1}^n g_i(w) X_i \ge \theta\}} - \mathbb{1}_{\{\sum_{i=1}^n g_i(\bar{w}) X_i \ge \theta\}} \right| \right] \le I_1 + I_2, \tag{49}$$

where

$$\begin{aligned}
I_1 &= \mathbb{E} \left[\sum_i |X_i| \left| \mathbb{1}_{\{\sum_{i=1}^n g_i(w) X_i \ge \theta\}} - \mathbb{1}_{\{\sum_{l \neq 1} g_l(w) X_l + g_1(\bar{w}) X_1 \ge \theta\}} \right| \right], \\
I_2 &= \mathbb{E} \left[\sum_i |X_i| \left| \mathbb{1}_{\{\sum_{l \neq 1} g_l(w) X_l + g_1(\bar{w}) X_1 \ge \theta\}} - \mathbb{1}_{\{\sum_{l \neq 1,2} g_l(w) X_l + g_1(\bar{w}) X_1 + g_2(\bar{w}) X_2 \ge \theta\}} \right| \right] \\
&\quad + \cdots \\
&\quad + \mathbb{E} \left[\sum_i |X_i| \left| \mathbb{1}_{\{g_n(w) X_n + \sum_{l \neq n} g_l(\bar{w}) X_l \ge \theta\}} - \mathbb{1}_{\{\sum_{i=1}^n g_i(\bar{w}) X_i \ge \theta\}} \right| \right].
\end{aligned}$$

To estimate I_1, one writes

$$
\begin{aligned}
I_1 \le{}& \mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta - \sum_{l\neq n} g_l(w)X_l)/g_n(w) \le X_n \le (\theta - g_1(\bar{w})X_1 - \sum_{l\neq 1,n} g_l(w)X_l)/g_n(w)\}}\right] \\
&+ \mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta - g_1(\bar{w})X_1 - \sum_{l\neq 1,n} g_l(w)X_l)/g_n(w)\} \le X_n \le (\theta - \sum_{l\neq n} g_l(w)X_l)/g_n(w)\}}\right].
\end{aligned}
$$

The first term on the RHS of the inequality above can be further estimated as

$$
\begin{aligned}
&\mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta - \sum_{l\neq n} g_l(w)X_l)/g_n(w) \le X_n \le (\theta - g_1(\bar{w})X_1 - \sum_{l\neq 1,n} g_l(w)X_l)/g_n(w)\}}\right] \\
&= \mathbb{E}\left[\sum_{i\neq n} |X_i| \mathbb{E}\left[\mathbb{1}_{\{(\theta - \sum_{l\neq n} g_l(w)X_l)/g_n(w) \le X_n \le (\theta - g_1(\bar{w})X_1 - \sum_{l\neq 1,n} g_l(w)X_l)/g_n(w)\}} \,\middle|\, X_1, \dots, X_{n-1}\right]\right] \\
&\quad + \mathbb{E}\left[\mathbb{E}\left[|X_n| \mathbb{1}_{\{(\theta - \sum_{l\neq n} g_l(w)X_l)/g_n(w) \le X_n \le (\theta - g_1(\bar{w})X_1 - \sum_{l\neq 1,n} g_l(w)X_l)/g_n(w)\}} \,\middle|\, X_1, \dots, X_{n-1}\right]\right] \\
&= \int_{-\infty}^{\infty} \sum_{i\neq n} |x_i| \cdots \int_{-\infty}^{\infty} \int_{(\theta - \sum_{l\neq n} g_l(w)x_l)/g_n(w)}^{(\theta - g_1(\bar{w})x_1 - \sum_{l\neq 1,n} g_l(w)x_l)/g_n(w)} f_{X_n}(z)\, dz \\
&\qquad\qquad\qquad \times f_{X_{n-1}}(x_{n-1})\, dx_{n-1} \cdots f_{X_1}(x_1)\, dx_1 \\
&\quad + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{(\theta - \sum_{l\neq n} g_l(w)x_l)/g_n(w)}^{(\theta - g_1(\bar{w})x_1 - \sum_{l\neq 1,n} g_l(w)x_l)/g_n(w)} |x_n| f_{X_n}(z)\, dz \\
&\qquad\qquad\qquad \times f_{X_{n-1}}(x_{n-1})\, dx_{n-1} \cdots f_{X_1}(x_1)\, dx_1 \\
&\le \frac{c_{X_n}(c_{\bar{X}} + (n-2)c_X^2)}{g_n(w)} |g_1(w) - g_1(\bar{w})| + \frac{\bar{c}_{X_n} c_X}{g_n(w)} |g_1(w) - g_1(\bar{w})| \\
&= (c_{X_n}(c_{\bar{X}} + (n-2)c_X^2) + \bar{c}_{X_n} c_X) \frac{\sum_i e^{w_i}}{e^{w_n}} \frac{\left(\sum_{i\neq 1} e^{w_i}\right) |e^{\bar{w}_1} - e^{w_1}|}{(\sum_i e^{w_i}) \left(e^{\bar{w}_1} + \sum_{i\neq 1} e^{w_i}\right)} \\
&\le (c_{X_n}(c_{\bar{X}} + (n-2)c_X^2) + \bar{c}_{X_n} c_X) \frac{\sum_{i\neq 1} g_i(w)}{g_n(w)} \frac{e^{\bar{w}_1}}{\left(e^{\bar{w}_1} + \sum_{i\neq 1} e^{w_i}\right)} |\bar{w}_1 - w_1| \\
&\le (c_{X_n}(c_{\bar{X}} + (n-2)c_X^2) + \bar{c}_{X_n} c_X)(n-1) \frac{e^{\bar{w}_1}}{\left(e^{\bar{w}_1} + \sum_{i\neq 1} e^{w_i}\right)} |\bar{w}_1 - w_1| \\
&\le (c_{X_n}(c_{\bar{X}} + (n-2)c_X^2) + \bar{c}_{X_n} c_X)(n-1) |\bar{w}_1 - w_1|,
\end{aligned}
$$

where $c_{\bar{X}}$ denotes the second absolute moment of X_i's , c_{X_n} is the upper bound of the density of X_n, and we use $1 - e^{-x} \le x$ for $x \ge 0$ in the third inequality. Moreover, I_2 can be upper bounded by

$$
\begin{aligned}
I_2 \le{}& \mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(w)X_l)/g_1(\bar{w}) \le X_1 \le (\theta - \sum_{l\neq 1,2} g_l(w)X_l - g_2(\bar{w})X_2)/g_1(\bar{w})\}}\right] \\
&+ \mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta - \sum_{l\neq 1,2} g_l(w)X_l - g_2(\bar{w})X_2)/g_1(\bar{w}) \le X_1 \le (\theta - \sum_{l\neq 1} g_l(w)X_l)/g_1(\bar{w})\}}\right] \\
&+ \cdots \\
&+ \mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta - g_n(w)X_n - \sum_{l\neq 1,n} g_l(\bar{w})X_l)/g_1(\bar{w}) \le X_1 \le (\theta - \sum_{l\neq 1} g_l(\bar{w})X_l)/g_1(\bar{w})\}}\right] \\
&+ \mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(\bar{w})X_l)/g_1(\bar{w}) \le X_1 \le (\theta - g_n(w)X_n - \sum_{l\neq 1,n} g_l(\bar{w})X_l)/g_1(\bar{w})\}}\right].
\end{aligned}
$$

The first term on the RHS of the inequality above can be calculated as

$$
\begin{aligned}
&\mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta-\sum_{l\neq 1} g_l(w)X_l)/g_1(\bar{w}) \le X_1 \le (\theta-\sum_{l\neq 1,2} g_l(w)X_l - g_2(\bar{w})X_2)/g_1(\bar{w})\}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[|X_1| \mathbb{1}_{\{(\theta-\sum_{l\neq 1} g_l(w)X_l)/g_1(\bar{w}) \le X_1 \le (\theta-\sum_{l\neq 1,2} g_l(w)X_l - g_2(\bar{w})X_2)/g_1(\bar{w})\}} \middle| X_2, \dots, X_n\right]\right] \\
&\quad + \mathbb{E}\left[\sum_{i\neq 1} |X_i| \mathbb{E}\left[\mathbb{1}_{\{(\theta-\sum_{l\neq 1} g_l(w)X_l)/g_1(\bar{w}) \le X_1 \le (\theta-\sum_{l\neq 1,2} g_l(w)X_l - g_2(\bar{w})X_2)/g_1(\bar{w})\}} \middle| X_2, \dots, X_n\right]\right] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{(\theta-\sum_{l\neq 1} g_l(w)x_l)/g_1(\bar{w})}^{(\theta-\sum_{l\neq 1,2} g_l(w)x_l - g_2(\bar{w})x_2)/g_1(\bar{w})} |z| f_{X_1}(z)\, dz f_{X_n}(x_n)\, dx_n \cdots f_{X_2}(x_2)\, dx_2 \\
&\quad + \int_{-\infty}^{\infty} \sum_{i\neq 1} |x_i| \cdots \int_{-\infty}^{\infty} \int_{(\theta-\sum_{l\neq 1} g_l(w)x_l)/g_1(\bar{w})}^{(\theta-\sum_{l\neq 1,2} g_l(w)x_l - g_2(\bar{w})x_2)/g_1(\bar{w})} f_{X_1}(z)\, dz \\
&\qquad\qquad \times f_{X_n}(x_n)\, dx_n \cdots f_{X_2}(x_2)\, dx_2 \\
&\le \frac{\bar{c}_{X_1} c_X}{g_1(\bar{w})} |g_2(w) - g_2(\bar{w})| + \frac{c_{X_1}(c_{\bar{X}} + (n-2)c_X^2)}{g_1(\bar{w})} |g_2(w) - g_2(\bar{w})| \\
&= (\bar{c}_{X_1} c_X + c_{X_1}(c_{\bar{X}} + (n-2)c_X^2)) \frac{\left(e^{\bar{w}_1} + \sum_{i\neq 1} e^{w_i}\right)}{e^{\bar{w}_1}} \frac{e^{w_2} |e^{\bar{w}_1} - e^{w_1}|}{(\sum_i e^{w_i}) \left(e^{\bar{w}_1} + \sum_{i\neq 1} e^{w_i}\right)} \\
&\le (\bar{c}_{X_1} c_X + c_{X_1}(c_{\bar{X}} + (n-2)c_X^2)) \frac{e^{w_2} e^{\bar{w}_1}}{e^{\bar{w}_1} (\sum_i e^{w_i})} |\bar{w}_1 - w_1| \\
&\le (\bar{c}_{X_1} c_X + c_{X_1}(c_{\bar{X}} + (n-2)c_X^2)) |\bar{w}_1 - w_1|,
\end{aligned}
$$

where c_X denotes the first absolute moment of X_i's and $\bar{c}_{X_1}$ is the upper bound of the function $|x| f_{X_1}$. Thus, in the case $\bar{w}_1 \ge w_1$, (49) becomes

$$
\begin{aligned}
&\mathbb{E}\left[\sum_i |X_i| \left| \mathbb{1}_{\{\sum_{i=1}^n g_i(w)X_i \ge \theta\}} - \mathbb{1}_{\{\sum_{i=1}^n g_i(\bar{w})X_i \ge \theta\}} \right|\right] \\
&\le 2(n-1)((c_{X_n} + c_{X_1})(c_{\bar{X}} + (n-2)c_X^2) + c_X(\bar{c}_{X_n} + \bar{c}_{X_1})) |\bar{w}_1 - w_1|.
\end{aligned}
$$

(ii) As for the case $w_1 > \bar{w}_1$, the calculations are close to the above, however, one considers a different splitting as follows

$$
\mathbb{E}\left[\sum_i |X_i| \left| \mathbb{1}_{\{\sum_{i=1}^n g_i(w)X_i \ge \theta\}} - \mathbb{1}_{\{\sum_{i=1}^n g_i(\bar{w})X_i \ge \theta\}} \right|\right] \le T_1 + T_2, \tag{50}
$$

where

$$
\begin{aligned}
T_1 &= \mathbb{E}\left[\sum_i |X_i| \left| \mathbb{1}_{\{\sum_{i=1}^n g_i(w)X_i \ge \theta\}} - \mathbb{1}_{\{\sum_{l\neq n} g_l(w)X_l + g_n(\bar{w})X_n \ge \theta\}} \right|\right] \\
&\quad + \cdots \\
&\quad + \mathbb{E}\left[\sum_i |X_i| \left| \mathbb{1}_{\{g_1(w)X_1 + g_2(w)X_2 + \sum_{l\neq 1,2} g_l(\bar{w})X_l \ge \theta\}} - \mathbb{1}_{\{g_1(w)X_1 + \sum_{l\neq 1} g_l(\bar{w})X_l \ge \theta\}} \right|\right], \\
T_2 &= \mathbb{E}\left[\sum_i |X_i| \left| \mathbb{1}_{\{g_1(w)X_1 + \sum_{l\neq 1} g_l(\bar{w})X_l \ge \theta\}} - \mathbb{1}_{\{\sum_{i=1}^n g_i(\bar{w})X_i \ge \theta\}} \right|\right].
\end{aligned}
$$

To estimate T_1, one calculates

$$
T_1 \le \mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta-\sum_{l\neq 1} g_l(w)X_l)/g_1(w) \le X_1 \le (\theta - g_n(\bar{w})X_n - \sum_{l\neq 1,n} g_l(w)X_l)/g_1(w)\}}\right]
$$

$$
\begin{aligned}
&+ \mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta - g_n(\bar{w})X_n - \sum_{l\neq 1,n} g_l(w)X_l)/g_1(w) \leq X_1 \leq (\theta - \sum_{l\neq 1} g_l(w)X_l)/g_1(w)\}}\right] \\
&+ \cdots \\
&+ \mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta - \sum_{l\neq 1,2} g_l(\bar{w})X_l - g_2(w)X_2)/g_1(w) \leq X_1 \leq (\theta - \sum_{l\neq 1} g_l(\bar{w})X_l)/g_1(w)\}}\right] \\
&+ \mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(\bar{w})X_l)/g_1(w) \leq X_1 \leq (\theta - \sum_{l\neq 1,2} g_l(\bar{w})X_l - g_2(w)X_2)/g_1(w)\}}\right].
\end{aligned}
$$

The first term on the RHS of the inequality above can be further calculated as

$$
\begin{aligned}
&\mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(w)X_l)/g_1(w) \leq X_1 \leq (\theta - g_n(\bar{w})X_n - \sum_{l\neq 1,n} g_l(w)X_l)/g_1(w)\}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[|X_1| \mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(w)X_l)/g_1(w) \leq X_1 \leq (\theta - g_n(\bar{w})X_n - \sum_{l\neq 1,n} g_l(w)X_l)/g_1(w)\}} \,\middle|\, X_2, \ldots, X_n\right]\right] \\
&\quad + \mathbb{E}\left[\sum_{i\neq 1} |X_i| \mathbb{E}\left[\mathbb{1}_{\{(\theta - \sum_{l\neq 1} g_l(w)X_l)/g_1(w) \leq X_1 \leq (\theta - g_n(\bar{w})X_n - \sum_{l\neq 1,n} g_l(w)X_l)/g_1(w)\}} \,\middle|\, X_2, \ldots, X_n\right]\right] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{(\theta - \sum_{l\neq 1} g_l(w)x_l)/g_1(w)}^{(\theta - g_n(\bar{w})x_n - \sum_{l\neq 1,n} g_l(w)x_l)/g_1(w)} |z| f_{X_1}(z)\, dz f_{X_n}(x_n)\, dx_n \cdots f_{X_2}(x_2)\, dx_2 \\
&\quad + \int_{-\infty}^{\infty} \sum_{i\neq 1} |x_i| \cdots \int_{-\infty}^{\infty} \int_{(\theta - \sum_{l\neq 1} g_l(w)x_l)/g_1(w)}^{(\theta - g_n(\bar{w})x_n - \sum_{l\neq 1,n} g_l(w)x_l)/g_1(w)} f_{X_1}(z)\, dz \\
&\qquad\qquad\qquad\qquad \times f_{X_n}(x_n)\, dx_n \cdots f_{X_2}(x_2)\, dx_2 \\
&\leq \frac{\bar{c}_{X_1} c_X}{g_1(w)} |g_n(w) - g_n(\bar{w})| + \frac{c_{X_1}(c_{\bar{X}} + (n-2)c_X^2)}{g_1(w)} |g_n(w) - g_n(\bar{w})| \\
&= (\bar{c}_{X_1} c_X + c_{X_1}(c_{\bar{X}} + (n-2)c_X^2)) \frac{\sum_i e^{w_i}}{e^{w_1}} \frac{e^{w_n} |e^{w_1} - e^{\bar{w}_1}|}{(\sum_i e^{w_i}) \left(e^{\bar{w}_1} + \sum_{i\neq 1} e^{w_i}\right)} \\
&\leq (\bar{c}_{X_1} c_X + c_{X_1}(c_{\bar{X}} + (n-2)c_X^2) \frac{e^{w_n} e^{w_1}}{e^{w_1} \left(e^{\bar{w}_1} + \sum_{i\neq 1} e^{w_i}\right)} |w_1 - \bar{w}_1| \\
&\leq (\bar{c}_{X_1} c_X + c_{X_1}(c_{\bar{X}} + (n-2)c_X^2)|w_1 - \bar{w}_1|.
\end{aligned}
$$

In addition, T_2 can be estimated as

$$
\begin{aligned}
T_2 \leq\ & \mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta - g_1(w)X_1 - \sum_{l\neq 1,n} g_l(\bar{w})X_l)/g_n(\bar{w}) \leq X_n \leq (\theta - \sum_{l\neq n} g_l(\bar{w})X_l)/g_n(\bar{w})\}}\right] \\
&+ \mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta - \sum_{l\neq n} g_l(\bar{w})X_l)/g_n(\bar{w}) \leq X_n \leq (\theta - g_1(w)X_1 - \sum_{l\neq 1,n} g_l(\bar{w})X_l)/g_n(\bar{w})\}}\right].
\end{aligned}
$$

The first term on the RHS of the above inequality can be upper bounded by

$$
\begin{aligned}
&\mathbb{E}\left[\sum_i |X_i| \mathbb{1}_{\{(\theta - g_1(w)X_1 - \sum_{l\neq 1,n} g_l(\bar{w})X_l)/g_n(\bar{w}) \leq X_n \leq (\theta - \sum_{l\neq n} g_l(\bar{w})X_l)/g_n(\bar{w})\}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[|X_n| \mathbb{1}_{\{(\theta - g_1(w)X_1 - \sum_{l\neq 1,n} g_l(\bar{w})X_l)/g_n(\bar{w}) \leq X_n \leq (\theta - \sum_{l\neq n} g_l(\bar{w})X_l)/g_n(\bar{w})\}} \,\middle|\, X_1, \ldots, X_{n-1}\right]\right] \\
&\quad + \mathbb{E}\left[\sum_{i\neq n} |X_i| \mathbb{E}\left[\mathbb{1}_{\{(\theta - g_1(w)X_1 - \sum_{l\neq 1,n} g_l(\bar{w})X_l)/g_n(\bar{w}) \leq X_n \leq (\theta - \sum_{l\neq n} g_l(\bar{w})X_l)/g_n(\bar{w})\}} \,\middle|\, X_1, \ldots, X_{n-1}\right]\right] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{(\theta - g_1(w)x_1 - \sum_{l\neq 1,n} g_l(\bar{w})x_l)/g_n(\bar{w})}^{(\theta - \sum_{l\neq n} g_l(\bar{w})x_l)/g_n(\bar{w})} |x_n| f_{X_n}(z)\, dz
\end{aligned}
$$

$$
\begin{aligned}
&\qquad\qquad\qquad\qquad \times f_{X_{n-1}}(x_{n-1})\,dx_{n-1}\cdots f_{X_1}(x_1)\,dx_1 \\
&+\int_{-\infty}^{\infty}\sum_{i\neq n}|x_i|\cdots\int_{-\infty}^{\infty}\int_{(\theta-g_1(w)x_1-\sum_{l\neq 1,n}g_l(\bar{w})x_l)/g_n(\bar{w})}^{(\theta-\sum_{l\neq n}g_l(\bar{w})x_l)/g_n(\bar{w})} f_{X_n}(z)\,dz \\
&\qquad\qquad\qquad\qquad \times f_{X_{n-1}}(x_{n-1})\,dx_{n-1}\cdots f_{X_1}(x_1)\,dx_1 \\
&\leq \frac{\bar{c}_{X_n}c_X}{g_n(\bar{w})}|g_1(w)-g_1(\bar{w})| + \frac{c_{X_n}(c_{\bar{X}}+(n-2)c_X^2)}{g_n(\bar{w})}|g_1(w)-g_1(\bar{w})| \\
&= (\bar{c}_{X_n}c_X + c_{X_n}(c_{\bar{X}}+(n-2)c_X^2))\frac{\left(e^{\bar{w}_1}+\sum_{i\neq 1}e^{w_i}\right)}{e^{w_n}}\frac{\left(\sum_{i\neq 1}e^{w_i}\right)|e^{w_1}-e^{\bar{w}_1}|}{\left(\sum_i e^{w_i}\right)\left(e^{\bar{w}_1}+\sum_{i\neq 1}e^{w_i}\right)} \\
&= (\bar{c}_{X_n}c_X + c_{X_n}(c_{\bar{X}}+(n-2)c_X^2))\frac{\sum_{i\neq 1}g_i(w)}{g_n(w)}\frac{e^{w_1}}{\sum_i e^{w_i}}|w_1-\bar{w}_1| \\
&\leq (n-1)(\bar{c}_{X_n}c_X + c_{X_n}(c_{\bar{X}}+(n-2)c_X^2))\frac{e^{w_1}}{\sum_i e^{w_i}}|w_1-\bar{w}_1| \\
&\leq (n-1)(\bar{c}_{X_n}c_X + c_{X_n}(c_{\bar{X}}+(n-2)c_X^2))|w_1-\bar{w}_1|.
\end{aligned}
$$

Thus for the case $w_1 > \bar{w}_1$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\sum_i |X_i|\left|\mathbb{1}_{\{\sum_{i=1}^n g_i(w)X_i\geq\theta\}}-\mathbb{1}_{\{\sum_{i=1}^n g_i(\bar{w})X_i\geq\theta\}}\right|\right] \\
&\leq 2(n-1)(c_X(\bar{c}_{X_n}+\bar{c}_{X_1}) + (c_{\bar{X}}+(n-2)c_X^2)(c_{X_n}+c_{X_1}))|w_1-\bar{w}_1|.
\end{aligned}
$$

Combining the two cases, one obtains

$$
\begin{aligned}
&\mathbb{E}\left[\sum_i |X_i|\left|\mathbb{1}_{\{\sum_{i=1}^n g_i(w)X_i\geq\theta\}}-\mathbb{1}_{\{\sum_{i=1}^n g_i(\bar{w})X_i\geq\theta\}}\right|\right] \\
&\leq 2(n-1)(c_X(\bar{c}_{X_n}+\bar{c}_{X_1}) + (c_{\bar{X}}+(n-2)c_X^2)(c_{X_n}+c_{X_1}))|w_1-\bar{w}_1|.
\end{aligned}
$$

A.4 Auxiliary results

Lemma 12. *Let Assumption 1, 2, 3 and 4 hold. For any $t\in[nT,(n+1)T]$, $n\in\mathbb{N}$ and $k=1,\ldots,K+1$, $K+1\leq T$, one obtains*

$$
\mathbb{E}\left[\left|H(\bar{\theta}^\lambda_{nT+k-1},X_{nT+k})-h(\bar{\theta}^\lambda_{nT+k-1})\right|^2\right]\leq e^{-a\lambda nT}\bar{\sigma}_Z\mathbb{E}[V_2(\theta_0)]+\tilde{\sigma}_Z,
$$

where

$$
\begin{aligned}
\bar{\sigma}_Z &= 4\mathbb{E}\left[K_\rho(X_0)\right]\left(L^2+L_1^2\right) \\
\tilde{\sigma}_Z &= 4\mathbb{E}\left[K_\rho(X_0)\right]\left(L^2+L_1^2\right)c_1(\lambda_{\max}+a^{-1})+4|h(0)|^2+8L_2^2\mathbb{E}\left[K_\rho(X_0)\right]+8\mathbb{E}\left[F_*^2(X_0)\right].
\end{aligned}
\tag{51}
$$

Proof. One notices that by Remark 1 and 2,

$$
\begin{aligned}
&\mathbb{E}\left[\left|H(\bar{\theta}^\lambda_{nT+k-1},X_{nT+k})-h(\bar{\theta}^\lambda_{nT+k-1})\right|^2\right] \\
&\leq 2\mathbb{E}\left[\left|h(\bar{\theta}^\lambda_{nT+k-1})\right|^2\right]+2\mathbb{E}\left[\left|H(\bar{\theta}^\lambda_{nT+k-1},X_{nT+k})\right|^2\right] \\
&\leq 2\mathbb{E}\left[\left(L\left|\bar{\theta}^\lambda_{nT+k-1}\right|+|h(0)|\right)^2\right]+2\mathbb{E}\left[\left((1+|X_{nT+k}|)^{\rho+1}\left(L_1\left|\bar{\theta}^\lambda_{nT+k-1}\right|+L_2\right)+F_*(X_{nT+k})\right)^2\right] \\
&\leq 4L^2\mathbb{E}\left[\left|\bar{\theta}^\lambda_{nT+k-1}\right|^2\right]+4|h(0)|^2+4L_1^2\mathbb{E}\left[K_\rho(X_0)\right]\mathbb{E}\left[\left|\bar{\theta}^\lambda_{nT+k-1}\right|^2\right]+8L_2^2\mathbb{E}\left[K_\rho(X_0)\right]+8\mathbb{E}\left[F_*^2(X_0)\right] \\
&\leq 4\mathbb{E}\left[K_\rho(X_0)\right]\left(L^2+L_1^2\right)\left(e^{-a\lambda nT}\mathbb{E}[V_2(\theta_0)]+c_1(\lambda_{\max}+a^{-1})\right)
\end{aligned}
$$

$$+ 4|h(0)|^2 + 8L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + 8\mathbb{E}\left[F_*^2(X_0)\right],$$

where the last inequality holds due to Lemma 1. Finally, one obtains

$$\mathbb{E}\left[\left|h(\bar{\zeta}_t^{\lambda,n}) - H(\bar{\zeta}_t^{\lambda,n}, X_{nT+k})\right|^2\right] \leq e^{-a\lambda nT}\bar{\sigma}_Z\mathbb{E}[V_2(\theta_0)] + \tilde{\sigma}_Z,$$

where $\bar{\sigma}_Z = 4\mathbb{E}\left[K_\rho(X_0)\right]\left(L^2 + L_1^2\right)$ and $\tilde{\sigma}_Z = 4\mathbb{E}\left[K_\rho(X_0)\right]\left(L^2 + L_1^2\right)c_1(\lambda_{\max} + a^{-1}) + 4|h(0)|^2 + 8L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + 8\mathbb{E}\left[F_*^2(X_0)\right]$. ◻

Lemma 13. *Let Assumption 1, 2 and 4 hold. For any $t > 0$, one obtains*

$$\mathbb{E}\left[\left|\bar{\theta}_t^\lambda - \bar{\theta}_{\lfloor t\rfloor}^\lambda\right|^2\right] \leq \lambda(e^{-a\lambda\lfloor t\rfloor}\bar{\sigma}_Y\mathbb{E}[V_2(\theta_0)] + \tilde{\sigma}_Y),$$

where

$$\begin{aligned}
&\bar{\sigma}_Y = 2\lambda_{\max}L_1^2\mathbb{E}\left[K_\rho(X_0)\right] \\
&\tilde{\sigma}_Y = 2\lambda_{\max}L_1^2\mathbb{E}\left[K_\rho(X_0)\right]c_1(\lambda_{\max} + a^{-1}) + 4\lambda_{\max}L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + 4\lambda_{\max}\mathbb{E}\left[F_*^2(X_0)\right] + 2d\beta^{-1}.
\end{aligned} \tag{52}$$

Proof. For any $t > 0$, one calculates

$$\begin{aligned}
\mathbb{E}\left[\left|\bar{\theta}_t^\lambda - \bar{\theta}_{\lfloor t\rfloor}^\lambda\right|^2\right] &= \mathbb{E}\left[\left|-\lambda\int_{\lfloor t\rfloor}^t H(\bar{\theta}_{\lfloor t\rfloor}^\lambda, X_{\lceil t\rceil})ds + \sqrt{2\beta^{-1}\lambda}(\tilde{B}_t^\lambda - \tilde{B}_{\lfloor t\rfloor}^\lambda)\right|^2\right] \\
&\leq \lambda^2\mathbb{E}\left[\left((1 + |X_{\lceil t\rceil}|)^{\rho+1}(L_1|\bar{\theta}_{\lfloor t\rfloor}^\lambda| + L_2) + F_*(X_{\lceil t\rceil})\right)^2\right] + 2d\lambda\beta^{-1},
\end{aligned}$$

where the inequality holds due to Remark 1 and by applying Lemma 1, one obtains

$$\begin{aligned}
\mathbb{E}\left[\left|\bar{\theta}_t^\lambda - \bar{\theta}_{\lfloor t\rfloor}^\lambda\right|^2\right] &\leq 2\lambda^2L_1^2\mathbb{E}\left[K_\rho(X_0)\right]\mathbb{E}[|\bar{\theta}_{\lfloor t\rfloor}^\lambda|^2] + 4\lambda^2L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + 4\lambda^2\mathbb{E}\left[F_*^2(X_0)\right] + 2d\lambda\beta^{-1} \\
&\leq \lambda((1 - a\lambda)^{\lfloor t\rfloor}\bar{\sigma}_Y\mathbb{E}[V_2(\theta_0)] + \tilde{\sigma}_Y),
\end{aligned}$$

where $\bar{\sigma}_Y = 2\lambda_{\max}L_1^2\mathbb{E}\left[K_\rho(X_0)\right]$ and $\tilde{\sigma}_Y = 2\lambda_{\max}L_1^2\mathbb{E}\left[K_\rho(X_0)\right]c_1(\lambda_{\max} + a^{-1}) + 4\lambda_{\max}L_2^2\mathbb{E}\left[K_\rho(X_0)\right] + 4\lambda_{\max}\mathbb{E}\left[F_*^2(X_0)\right] + 2d\beta^{-1}$. ◻

Lemma 14. *Let Assumption 1, 2, and 4 hold. Then, for any $t > 0$, one obtains*

$$\mathbb{E}[|Z_t|^2] \leq e^{-at}\mathbb{E}[|\theta_0|^2] + \left(\frac{2d}{a\beta} + \frac{2b}{a} + \frac{\mathbb{E}[K_1^2(X_0)]}{a^2}\right)(1 - e^{-at}).$$

Proof. For any $t > 0$, by applying Itô's formula to $e^{at}|Z_t|^2$, one obtains, almost surely

$$de^{at}|Z_t|^2 = ae^{at}|Z_t|^2dt - 2e^{at}\langle Z_t, h(Z_t)\rangle dt + 2e^{at}\langle Z_t, \sqrt{2\beta^{-1}}dB_t\rangle + 2d\beta^{-1}e^{at}dt.$$

Then, integrating both sides and taking expectation yield

$$e^{at}\mathbb{E}[|Z_t|^2] = \mathbb{E}[|\theta_0|^2] + a\int_0^t e^{as}\mathbb{E}[|Z_s|^2]ds - 2\int_0^t e^{as}\mathbb{E}[\langle Z_s, h(Z_s)\rangle]ds + 2d\beta^{-1}\int_0^t e^{as}ds,$$

which implies by using Assumption 4

$$\begin{aligned}
e^{at}\mathbb{E}[|Z_t|^2] &= \mathbb{E}[|\theta_0|^2] + a\int_0^t e^{as}\mathbb{E}[|Z_s|^2]ds - 2a\int_0^t e^{as}\mathbb{E}[|Z_s|^2]ds + 2b\int_0^t e^{as}ds \\
&\quad + 2\int_0^t e^{as}\mathbb{E}[|Z_s|]\mathbb{E}[K_1(X_0)]ds + 2d\beta^{-1}\int_0^t e^{as}ds \\
&\leq \mathbb{E}[|\theta_0|^2] + (2b + \mathbb{E}[K_1^2(X_0)]/a + 2d\beta^{-1})(e^{at} - 1)/a.
\end{aligned}$$

Finally, one obtains

$$\mathbb{E}[|Z_t|^2] \leq e^{-at}\mathbb{E}[|\theta_0|^2] + (2b + \mathbb{E}[K_1^2(X_0)]/a + 2d\beta^{-1})(1 - e^{-at})/a.$$

◻